

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2002, 11:15:22 ; Search time 30.05 Seconds  
(without alignments)  
66.533 Million cell updates/sec

Title: US-09-394-019A-248  
Perfect score: 96  
Sequence: 1 KDPXGLEHNGINGXPKGY 18

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	95.8	18	22	Protease indicator
2	90	93.8	18	22	Protease binding s
3	86	89.6	18	22	Protease indicator
4	84	87.5	18	22	Protease binding s
5	82	85.4	18	22	Protease indicator
6	80	83.3	18	22	Protease binding s
7	80	83.3	18	22	Protease indicator
8	78	81.2	18	22	Protease binding s
9	73	76.0	18	19	Fluorogenic protea
10	73	76.0	18	22	Protease indicator
11	73	76.0	18	22	Protease indicator

12	71	74.0	18	22	AAG73082	Protease binding s
13	71	74.0	18	22	AAG73086	Protease binding s
14	71	74.0	18	22	AAG73090	Protease binding s
15	71	74.0	18	22	AAG73091	Protease binding s
16	70	72.9	18	22	AAG73318	Protease indicator
17	68	70.8	18	22	AAG73088	Protease binding s
18	65.5	68.2	19	22	AAG73302	Protease indicator
19	65	67.7	18	19	AAW82120	Fluorogenic protea
20	65	67.7	18	22	AAG73319	Protease indicator
21	64	66.7	18	19	AAW82119	Fluorogenic protea
22	64	66.7	18	19	AAW82115	Fluorogenic protea
23	63.5	66.1	17	22	AAG73303	Protease indicator
24	63.5	66.1	17	22	AAG73304	Protease binding s
25	63.5	66.1	19	22	AAG73084	Protease binding s
26	63	65.6	18	22	AAG73089	Protease binding s
27	63	65.6	18	22	AAG73099	Protease binding s
28	62.5	65.1	17	22	AAG73085	Protease binding s
29	62	64.6	18	22	AAG73094	Protease binding s
30	62	64.6	18	22	AAG73098	Protease binding s
31	61	63.5	18	19	AAW82136	Protease indicator
32	60	62.5	16	22	AAG73301	Protease indicator
33	60	62.5	18	19	AAW82118	Fluorogenic protea
34	60	62.5	18	19	AAW82122	Fluorogenic protea
35	59	61.5	18	19	AAW82121	Fluorogenic protea
36	59	61.5	18	19	AAW82110	Fluorogenic protea
37	59	61.5	18	19	AAW82113	Fluorogenic protea
38	59	61.5	18	19	AAW82114	Fluorogenic protea
39	59	61.5	18	19	AAW82238	Fluorogenic protea
40	59	61.5	18	19	AAW82240	Fluorogenic protea
41	59	61.5	18	22	AAG73079	Protease binding s
42	59	61.5	18	22	AAG73080	Protease binding s
43	59	61.5	18	22	AAG73117	Protease binding s
44	59	61.5	18	22	AAG73118	Protease binding s
45	59	61.5	18	22	AAG73258	Protease indicator

ALIGNMENTS

RESULT 1  
AAG73324  
ID AAG73324 standard; Peptide; 18 AA.  
XX AC AAG73324;  
XX AC AAG73324;  
XX DT 14-AUG-2001 (first entry)  
XX DE Protease indicator compound peptide #53.  
XX DE Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.  
XX OS Synthetic.  
XX FH Key  
XX FT Modified-site 4  
XX FT /label= OTHER  
XX FT /note= "designated J in the specification"  
XX FT Modified-site 14  
XX FT /label= OTHER  
XX FT /note= "designated J in the specification"

WO200118238-A1.

15-MAR-2001.

11-SEP-2000; 2000WO-US24882.

10-SEP-1999; 99US-0394019.

(ONCO-) ONCOIMMUNIN INC.



CC detected. The present sequence is one of the peptides described in the  
XX exemplification of the invention.  
SQ Sequence 18 AA;

Query Match 89.6%; Score 86; DB 22; Length 18;  
Best Local Similarity 94.4%; Pred. No. 3.9e-08;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
| | | | | | | | | | | | | | | | | |  
Db 1 kdpxgwehdgngxpkgy 18

RESULT 4  
AAG73232  
ID AAG73232 standard; Peptide; 18 AA.  
XX  
AC AAG73232;  
XX  
DT 14-AUG-2001 (first entry)  
XX  
DE Protease binding site #166.  
XX  
KW Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.  
XX  
OS Synthetic.  
XX

Key Location/Qualifiers  
FH Modified-site 1  
FT /label= OTHER  
FT /note= "modified by fluorophore"  
XX  
PN WO200118238-A1.  
XX  
PD 15-MAR-2001.  
XX

PF 11-SEP-2000; 2000WO-US24882.  
XX  
PR 10-SEP-1999; 99US-0394019.  
XX  
PA (ONCO-) ONCOIMMUNIN INC.  
XX  
PI Komoriya A, Packard BS;  
XX  
DR WPI; 2001-389573/41.  
XX

PT New fluorogenic compositions whose fluorescence level increases in the  
PT presence of active proteases, useful for detecting and localizing  
PT protease activity in biological samples, particularly in frozen tissue  
PT samples -  
XX  
PS Disclosure; Page 29; 86pp; English.

CC The present invention describes fluorogenic compositions which can be  
CC used for the detection of protease activity. This can be useful as an  
CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
CC thrombosis and arthritis. The fluorogenic compositions comprise a  
CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
CC peptide is cleaved by a protease and the fluorophores can then be  
CC detected. The present sequence is one of the peptides described in the  
CC exemplification of the invention.

SQ Sequence 18 AA;

Query Match 87.5%; Score 84; DB 22; Length 18;  
Best Local Similarity 83.3%; Pred. No. 8.7e-08;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
| | | | | | | | | | | | | | | | | |  
Db 1 kdpcgwehdgngcpgky 18

RESULT 5  
AAG73325  
ID AAG73325 standard; Peptide; 18 AA.  
XX  
AC AAG73325;  
XX  
DT 14-AUG-2001 (first entry)  
XX  
DE Protease indicator compound peptide #54.  
XX

KW Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.  
XX  
OS Synthetic.  
XX

Key Location/Qualifiers  
FH Modified-site 4  
FT /label= OTHER  
FT /note= "designated J in the specification"  
FT Modified-site 14  
FT /label= OTHER  
FT /note= "designated J in the specification"  
XX  
PN WO200118238-A1.  
XX  
PD 15-MAR-2001.  
XX

PF 11-SEP-2000; 2000WO-US24882.  
XX  
PR 10-SEP-1999; 99US-0394019.  
XX  
PA (ONCO-) ONCOIMMUNIN INC.  
XX  
PI Komoriya A, Packard BS;  
XX  
DR WPI; 2001-389573/41.  
XX

PT New fluorogenic compositions whose fluorescence level increases in the  
PT presence of active proteases, useful for detecting and localizing  
PT protease activity in biological samples, particularly in frozen tissue  
PT samples -  
XX

PS Claim 4; Page 71; 86pp; English.

CC The present invention describes fluorogenic compositions which can be  
CC used for the detection of protease activity. This can be useful as an  
CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
CC thrombosis and arthritis. The fluorogenic compositions comprise a  
CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
CC peptide is cleaved by a protease and the fluorophores can then be  
CC detected. The present sequence is one of the peptides described in the  
CC exemplification of the invention.

SQ Sequence 18 AA;

Query Match 85.4%; Score 82; DB 22; Length 18;  
Best Local Similarity 94.4%; Pred. No. 1.9e-07;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
| | | | | | | | | | | | | | | | | |  
Db 1 kdpxgletdngxpkgy 18

RESULT 6  
AAG73231

Tue Jul 16 16:10:13 2002

thrombosis; haemophilia.  
Synthetic.  
Key Location/Qualifiers  
Modified-site 4 /label= OTHER  
Modified-site 14 /note= "designated J in the specification"  
Modified-site 14 /label= OTHER  
Modified-site 14 /note= "designated J in the specification"  
WO200118238-A1.  
15-MAR-2001.  
11-SEP-2000; 2000WO-US24882.  
10-SEP-1999; 99US-0394019.  
(ONCO-) ONCOIMMUNIN INC.  
Komoriya A, Packard BS;  
WPI; 2001-389573/41.  
New fluorogenic compositions whose fluorescence level increases in the presence of active proteases, useful for detecting and localizing protease activity in biological samples, particularly in frozen tissue samples  
Claim 4; Page 71; 86pp; English.  
The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention.  
Sequence 18 AA;  
Query Match 83.3%; Score 80; DB 22; Length 18;  
Best Local Similarity 88.9%; Pred. No. 4.3e-07;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 KDPXGLEHGDINGXPKGY 18  
Db 1 kdpxyvhdgngxpkgy 18  
RESULT 8  
AAG73234  
ID AAG73234 standard; Peptide; 18 AA.  
XX AC AAG73234;  
XX DT 14-AUG-2001 (first entry)  
XX DE Protease binding site #165.  
XX DE Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
XX DE viral infection; cancer metastasis; emphysema; arthritis;  
XX DE thrombosis; haemophilia.  
XX OS Synthetic.  
XX FH Key Location/Qualifiers  
XX FT Modified-site 1 /label= OTHER  
XX FT /note= "modified by fluorophore"  
WO200118238-A1.  
15-MAR-2001.  
11-SEP-2000; 2000WO-US24882.  
10-SEP-1999; 99US-0394019.  
(ONCO-) ONCOIMMUNIN INC.  
Komoriya A, Packard BS;  
WPI; 2001-389573/41.  
New fluorogenic compositions whose fluorescence level increases in the presence of active proteases, useful for detecting and localizing protease activity in biological samples, particularly in frozen tissue samples  
Disclosure; Page 29; 86pp; English.  
The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention.  
Sequence 18 AA;  
Query Match 83.3%; Score 80; DB 22; Length 18;  
Best Local Similarity 83.3%; Pred. No. 4.3e-07;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 KDPXGLEHGDINGXPKGY 18  
Db 1 kdpcegtgdngcpgky 18  
RESULT 7  
AAG73328  
ID AAG73328 standard; Peptide; 18 AA.  
XX AC AAG73328;  
XX DT 14-AUG-2001 (first entry)  
XX DE Protease indicator compound peptide #57.  
XX DE Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
XX DE viral infection; cancer metastasis; emphysema; arthritis;  
KW

thrombosis; haemophilia.  
Synthetic.  
Key Location/Qualifiers  
Modified-site 4 /label= OTHER  
Modified-site 14 /note= "designated J in the specification"  
Modified-site 14 /label= OTHER  
Modified-site 14 /note= "designated J in the specification"  
WO200118238-A1.  
15-MAR-2001.  
11-SEP-2000; 2000WO-US24882.  
10-SEP-1999; 99US-0394019.  
(ONCO-) ONCOIMMUNIN INC.  
Komoriya A, Packard BS;  
WPI; 2001-389573/41.  
New fluorogenic compositions whose fluorescence level increases in the presence of active proteases, useful for detecting and localizing protease activity in biological samples, particularly in frozen tissue samples  
Claim 4; Page 71; 86pp; English.  
The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention.  
Sequence 18 AA;  
Query Match 83.3%; Score 80; DB 22; Length 18;  
Best Local Similarity 88.9%; Pred. No. 4.3e-07;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 KDPXGLEHGDINGXPKGY 18  
Db 1 kdpxyvhdgngxpkgy 18  
RESULT 8  
AAG73234  
ID AAG73234 standard; Peptide; 18 AA.  
XX AC AAG73234;  
XX DT 14-AUG-2001 (first entry)  
XX DE Protease binding site #168.  
XX DE Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
XX DE viral infection; cancer metastasis; emphysema; arthritis;  
XX DE thrombosis; haemophilia.  
XX OS Synthetic.  
XX FH Key Location/Qualifiers  
XX FT Modified-site 1 /label= OTHER  
FT

FT XX /note= "modified by fluorophore"

PN WO200118238-A1.

XX 15-MAR-2001.

XX 11-SEP-2000; 2000WO-US24882.

PF 10-SEP-1999; 99US-0394019.

PR (ONCO-) ONCOIMMUNIN INC.

XX Komoriya A, Packard BS;

PI WPI; 2001-389573/41.

DR New fluorogenic compositions whose fluorescence level increases in the presence of active proteases, useful for detecting and localizing protease activity in biological samples, particularly in frozen tissue samples.

XX Disclosure; Page 29; 86pp; English.

CC The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention.

CC Sequence 18 AA;

XX

SQ

Query Match 81.2%; Score 78; DB 22; Length 18;

Best Local Similarity 77.8%; Pred. No. 9.5e-07;

Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18

DB 1 KDPXGYYHGDINGCPKGY 18

RESULT 9

AAW82112

ID AAW82112 standard; peptide; 18 AA.

XX AAW82112;

XX 18-FEB-1999 (first entry)

DE Fluorogenic protease indicator CPP32 substrate #1.

XX Protease activity; fluorophore; detection; fluorogenic; cellular uptake;

KW conformation change.

XX Synthetic.

OS

FH Key Location/Qualifiers

FT Modified-site 1

FT /note= "Lys residue modified by presence of Fmoc (9-fluorenylmethoxycarbonyl)"

FT Modified-site 4

FT /note= "epsilon-aminocaproic acid, labelled as amino acid J in the specification"

FT Modified-site 14

FT /note= "epsilon-aminocaproic acid, labelled as amino acid J in the specification"

WO9837226-A1.

XX 27-AUG-1998.

XX

XX 20-FEB-1998; 98WO-US03000.

XX 20-FEB-1997; 97US-0802981.

PR (ONCO-) ONCOIMMUNIN INC.

XX Komoriya A, Packard BS;

PI WPI; 1998-467579/40.

DR New fluorogenic compositions - containing 2 fluorophores separated by a peptide comprising a protease binding site, used for detecting protease activity in samples.

XX Disclosure; Page 24; 90pp; English.

XX AAW82023-W82240 are peptides used in the construction of a fluorogenic composition which is used for the detection of protease activity in biological samples. The products can be used for the detection of conformation changes in nucleic acids, oligosaccharides, polysaccharides, proteins, peptides, lipids, phospholipids, glycoproteins, steroids or polymers. In addition, attachment of a hydrophobic group to a molecule can be used to enhance uptake by cells. The composition is composed of P - peptide comprising a protease binding site for the protease, F1, F2 peptides - fluorophores where F1 is attached to the amino terminal amino acid and F2 is attached to the carboxyl terminal amino acid and S1, S2 peptides - when present, are peptide spacers where S1, when present, is attached to the amino terminal acid, and S2, when present, is attached to the carboxyl terminal amino acid.

XX Sequence 18 AA;

SQ

Query Match 76.0%; Score 73; DB 19; Length 18;

Best Local Similarity 88.9%; Pred. No. 7e-06;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18

DB 1 KDPXGDEVGDINGXPKGY 18

RESULT 10

AAG73300

ID AAG73300 standard; Peptide; 18 AA.

XX AAG73300;

XX 14-AUG-2001 (first entry)

DE Protease indicator compound peptide #29.

XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;

KW viral infection; cancer metastasis; emphysema; arthritis;

XX thrombosis; haemophilia.

OS Synthetic.

FH Key Location/Qualifiers

FT Modified-site 1

FT /label= OTHER

FT /note= "modified by Fa"

FT Modified-site 4

FT /label= OTHER

FT /note= "designated J in the specification"

FT Modified-site 14

FT /label= OTHER

FT /note= "designated J in the specification"

WO200118238-A1.

XX

us-09-394-019a-248.rag

Tue Jul 16 16:10:13 2002

```

PD 15-MAR-2001.
XX
XX 11-SEP-2000; 2000WO-US24882.
XX
XX 10-SEP-1999; 99US-0394019.
XX
XX (ONCO-) ONCOIMMUNIN INC.
XX
XX Komoriya A, Packard BS;
XX
XX WPI; 2001-389573/41.
XX
XX New fluorogenic compositions whose fluorescence level increases in the
XX presence of active proteases, useful for detecting and localizing
XX protease activity in biological samples, particularly in frozen tissue
XX samples.
XX
XX Claim 4; Page 71; 86pp; English.
XX
XX The present invention describes fluorogenic compositions which can be
XX used for the detection of protease activity. This can be useful as an
XX indicator of viral infection, cancer metastasis, haemophilia, emphysema,
XX thrombosis and arthritis. The fluorogenic compositions comprise a
XX peptide, a peptide spacer and a donor and an acceptor fluorophore. The
XX peptide is cleaved by a protease and the fluorophores can then be
XX detected. The present sequence is one of the peptides described in the
XX exemplification of the invention.
XX
XX Sequence 18 AA;
XX
XX Query Match 76.0%; Score 73; DB 22; Length 18;
XX Best Local Similarity 88.9%; Pred. No. 7e-06;
XX Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KDPXGLEHDGNGXPKGY 18
DB 1 kdpngxdevdngxpkgy 18

RESULT 11
AAG73320
ID AAG73320 standard; Peptide; 18 AA.
XX
XX AAG73320;
XX
XX 14-AUG-2001 (first entry)
XX
XX Protease indicator compound peptide #49.
XX
XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;
XX viral infection; cancer metastasis; emphysema; arthritis;
XX thrombosis; haemophilia.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 4 /label= OTHER
XX Misc-difference 6 /note= "designated J in the specification"
XX Misc-difference 9 /note= "D-form residue"
XX Misc-difference 14 /note= "D-form residue"
XX Modified-site 14 /label= OTHER
XX /note= "designated J in the specification"
XX
XX WO200118238-A1.
XX
XX 15-MAR-2001.
XX
XX 11-SEP-2000; 2000WO-US24882.
XX
XX 10-SEP-1999; 99US-0394019.
XX
XX (ONCO-) ONCOIMMUNIN INC.
XX
XX Komoriya A, Packard BS;
XX
XX WPI; 2001-389573/41.
XX
XX New fluorogenic compositions whose fluorescence level increases in the
XX presence of active proteases, useful for detecting and localizing
XX protease activity in biological samples, particularly in frozen tissue
XX samples.
XX
XX Claim 4; Page 71; 86pp; English.
XX
XX The present invention describes fluorogenic compositions which can be
XX used for the detection of protease activity. This can be useful as an
XX indicator of viral infection, cancer metastasis, haemophilia, emphysema,
XX thrombosis and arthritis. The fluorogenic compositions comprise a
XX peptide, a peptide spacer and a donor and an acceptor fluorophore. The
XX peptide is cleaved by a protease and the fluorophores can then be
XX detected. The present sequence is one of the peptides described in the
XX exemplification of the invention.
XX
XX Sequence 18 AA;
XX
XX Query Match 76.0%; Score 73; DB 22; Length 18;
XX Best Local Similarity 88.9%; Pred. No. 7e-06;
XX Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KDPXGLEHDGNGXPKGY 18
DB 1 kdpngxdevdngxpkgy 18

RESULT 12
AAG73082
ID AAG73082 standard; Peptide; 18 AA.
XX
XX AAG73082;
XX
XX 14-AUG-2001 (first entry)
XX
XX Protease binding site #16.
XX
XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;
XX viral infection; cancer metastasis; emphysema; arthritis;
XX thrombosis; haemophilia.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /label= OTHER
XX /note= "modified by Fa"
XX
XX WO200118238-A1.
XX
XX 15-MAR-2001.
XX
XX 11-SEP-2000; 2000WO-US24882.
XX
XX 10-SEP-1999; 99US-0394019.
XX
XX (ONCO-) ONCOIMMUNIN INC.
XX
XX Komoriya A, Packard BS;
XX
XX WPI; 2001-389573/41.
XX
XX New fluorogenic compositions whose fluorescence level increases in the
XX

```

PT Presence of active proteases, useful for detecting and localizing  
PT protease activity in biological samples, particularly in frozen tissue  
XX samples -  
PS Disclosure; Page 23; 86pp; English.  
XX  
CC The present invention describes fluorogenic compositions which can be  
CC used for the detection of protease activity. This can be useful as an  
CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
CC thrombosis and arthritis. The fluorogenic compositions comprise a  
CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
CC peptide is cleaved by a protease and the fluorophores can then be  
CC detected. The present sequence is one of the peptides described in the  
CC exemplification of the invention.  
XX  
SQ Sequence 18 AA;

Query Match 74.0%; Score 71; DB 22; Length 18;  
Best Local Similarity 77.8%; Pred. No. 1.5e-05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 KDPXGLEHDGNGXPKGY 18  
| | | | | | | | | | | | | | | |  
DB 1 kdpcgdevdgingcpkgy 18

RESULT 13  
AAG73086  
ID AAG73086 standard; Peptide; 18 AA.  
AC AAG73086;  
DT 14-AUG-2001 (first entry)  
DE Protease binding site #20.  
XX  
XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 1  
FT /label= OTHER  
FT /note= "modified by fluorophore"  
XX  
XX WO200118238-A1.  
XX  
XX 15-MAR-2001.  
XX  
XX 11-SEP-2000; 2000WO-US24882.  
XX  
XX 10-SEP-1999; 99US-0394019.  
XX  
XX (ONCO-) ONCOIMMUNIN INC.  
XX  
XX Komoriya A, Packard BS;  
XX  
XX WPI; 2001-389573/41.  
XX  
XX New fluorogenic compositions whose fluorescence level increases in the  
XX presence of active proteases, useful for detecting and localizing  
XX protease activity in frozen tissue  
XX samples -  
XX  
XX Disclosure; Page 23; 86pp; English.  
XX  
XX The present invention describes fluorogenic compositions which can be  
XX used for the detection of protease activity. This can be useful as an  
XX indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
XX thrombosis and arthritis. The fluorogenic compositions comprise a

CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
CC peptide is cleaved by a protease and the fluorophores can then be  
CC detected. The present sequence is one of the peptides described in the  
XX exemplification of the invention.  
SQ Sequence 18 AA;

Query Match 74.0%; Score 71; DB 22; Length 18;  
Best Local Similarity 77.8%; Pred. No. 1.5e-05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 KDPXGLEHDGNGXPKGY 18  
| | | | | | | | | | | | | | | |  
DB 1 kdpcgdevdgingcpkgy 18

RESULT 14  
AAG73090  
ID AAG73090 standard; Peptide; 18 AA.  
XX  
XX AAG73090;  
DT 14-AUG-2001 (first entry)  
DE Protease binding site #24.  
XX  
XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 1  
FT /label= OTHER  
FT /note= "modified by fluorophore"  
XX  
XX Misc-difference 6  
FT /note= "D-form residue"  
XX  
XX Misc-difference 9  
FT /note= "D-form residue"  
XX  
XX WO200118238-A1.  
XX  
XX 15-MAR-2001.  
XX  
XX 11-SEP-2000; 2000WO-US24882.  
XX  
XX 10-SEP-1999; 99US-0394019.  
XX  
XX (ONCO-) ONCOIMMUNIN INC.  
XX  
XX Komoriya A, Packard BS;  
XX  
XX WPI; 2001-389573/41.  
XX  
XX New fluorogenic compositions whose fluorescence level increases in the  
XX presence of active proteases, useful for detecting and localizing  
XX protease activity in biological samples, particularly in frozen tissue  
XX samples -  
XX  
XX Disclosure; Page 23; 86pp; English.  
XX  
XX The present invention describes fluorogenic compositions which can be  
XX used for the detection of protease activity. This can be useful as an  
XX indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
XX thrombosis and arthritis. The fluorogenic compositions comprise a  
XX peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
XX peptide is cleaved by a protease and the fluorophores can then be  
XX detected. The present sequence is one of the peptides described in the  
XX exemplification of the invention.  
XX  
XX Sequence 18 AA;

Tue Jul 16 16:10:13 2002

Search completed: July 16, 2002, 11:21:27  
Job time: 365 sec

Query Match 74.0%; Score 71; DB 22; Length 18;  
Best Local Similarity 77.8%; Pred. No. 1.5e-05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDGNGXPKGY 18  
||| | | |||| | |||  
DB 1 kdpdgdevdgingcpkgy 18

RESULT 15  
AAG73091  
ID AAG73091 standard; Peptide; 18 AA.  
XX  
AC AAG73091;  
DT 14-AUG-2001 (first entry)  
XX  
DE Protease binding site #25.  
XX  
KW Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 1  
FT /label= OTHER  
FT /note= "modified by fluorophore"  
XX  
PN WO200118238-A1.  
XX  
PD 15-MAR-2001.  
XX  
XX 11-SEP-2000; 2000WO-US24882.  
PF  
XX 10-SEP-1999; 99US-0394019.  
PR  
XX (ONCO-) ONCOIMMUNIN INC.  
PA  
XX Komoriya A, Packard BS;  
PI  
XX WPI; 2001-389573/41.  
DR  
XX New fluorogenic compositions whose fluorescence level increases in the  
PT presence of active proteases, useful for detecting and localizing  
PT protease activity in biological samples, particularly in frozen tissue  
PT samples -  
PT  
XX Disclosure; Page 23; 86pp; English.  
PS  
XX The present invention describes fluorogenic compositions which can be  
CC used for the detection of protease activity. This can be useful as an  
CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
CC thrombosis and arthritis. The fluorogenic compositions comprise a  
CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
CC peptide is cleaved by a protease and the fluorophores can then be  
CC detected. The present sequence is one of the peptides described in the  
CC exemplification of the invention.  
XX  
SQ Sequence 18 AA;

Query Match 74.0%; Score 71; DB 22; Length 18;  
Best Local Similarity 77.8%; Pred. No. 1.5e-05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDGNGXPKGY 18  
||| | | |||| | |||  
DB 1 kdpdgdevdgingcpkgy 18



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2002, 11:15:22 ; Search time 12.98 Seconds  
(without alignments)  
33.872 Million cell updates/sec

Title: US-09-394-019a-248  
Perfect score: 96  
Sequence: 1 KDPXGLEHGDINGXPKGY 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	65	67.7		18	3	US-08-802-981-55	Sequence 55, Appl
2	64	66.7		18	3	US-08-802-981-54	Sequence 54, Appl
3	61	63.5		18	3	US-08-802-981-69	Sequence 53, Appl
4	60	62.5		18	3	US-08-802-981-53	Sequence 53, Appl
5	60	62.5		18	3	US-08-802-981-57	Sequence 57, Appl
6	59	61.5		18	3	US-08-802-981-56	Sequence 56, Appl
7	56	58.3		18	3	US-08-802-981-66	Sequence 66, Appl
8	56	58.3		18	3	US-08-802-981-68	Sequence 68, Appl
9	54	56.2		18	3	US-08-802-981-49	Sequence 49, Appl
10	54	56.2		18	3	US-08-802-981-50	Sequence 50, Appl
11	51	53.1		18	3	US-08-802-981-67	Sequence 67, Appl
12	50.5	52.6		19	3	US-08-802-981-59	Sequence 59, Appl
13	48	50.0		18	3	US-08-802-981-173	Sequence 173, App
14	48	50.0		18	3	US-08-802-981-176	Sequence 174, App
15	47.5	49.5		19	3	US-08-802-981-58	Sequence 58, Appl
16	44.5	46.4		19	3	US-08-802-981-52	Sequence 52, Appl
17	44	45.8		19	3	US-08-802-981-64	Sequence 64, Appl
18	43.5	45.3		16	3	US-08-802-981-52	Sequence 48, Appl
19	43.5	45.3		655	4	US-03-245-808-1	Sequence 1, Appl
20	41.5	43.2		16	3	US-08-802-981-48	Sequence 48, Appl
21	40	41.7		18	3	US-08-802-981-73	Sequence 73, Appl
22	40	41.7		19	3	US-08-802-981-80	Sequence 80, Appl
23	39	40.6		18	3	US-08-802-981-72	Sequence 72, Appl
24	39	40.6		232	4	US-08-979-608A-3	Sequence 3, Appl
25	39	40.6		252	4	US-08-979-608A-4	Sequence 4, Appl
26	39	40.6		317	4	US-08-979-608A-2	Sequence 2, Appl
27	38.5	40.1		16	3	US-08-802-981-63	Sequence 63, Appl

28	38.5	40.1	16	3	US-08-802-981-65	Sequence 65, Appl
29	38	39.6	16	3	US-08-802-981-71	Sequence 71, Appl
30	38	39.6	19	3	US-08-802-981-86	Sequence 86, Appl
31	38	39.6	1297	4	US-09-540-245A-17	Sequence 17, Appl
32	38	39.6	1461	2	US-08-993-228-10	Sequence 10, Appl
33	37.5	39.1	383	4	US-09-413-814-12	Sequence 12, Appl
34	37	38.5	229	3	US-09-040-285A-7	Sequence 7, Appl
35	37	38.5	2556	1	US-08-185-432-17	Sequence 17, Appl
36	37	38.5	2556	3	US-08-083-590A-20	Sequence 20, Appl
37	37	38.5	2556	3	US-08-532-981-20	Sequence 20, Appl
38	36.5	38.0	17	3	US-08-802-981-78	Sequence 61, Appl
39	36	37.5	14	3	US-08-802-981-61	Sequence 38, Appl
40	36	37.5	43	2	US-08-488-161-38	Sequence 38, Appl
41	36	37.5	43	3	US-09-273-685-38	Sequence 38, Appl
42	36	37.5	43	5	PCT-US95-11934-38	Sequence 11, Appl
43	36	37.5	185	4	US-08-975-762-11	Sequence 11, Appl
44	36	37.5	185	4	US-08-821-324-11	Sequence 11, Appl
45	36	37.5	185	4	US-09-295-028-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-08-802-981-55  
; Sequence 55, Application US/08802981  
; Patent No. 6037137  
; GENERAL INFORMATION:  
; APPLICANT: Komoriya, Akira  
; APPLICANT: Packard, Beverly S.  
; TITLE OF INVENTION: Compositions for the Detection of Enzyme  
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 231  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/802,981  
; FILING DATE: 20-FEB-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 016865-00030005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 55:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 3  
; OTHER INFORMATION: /product= "Aib"  
; NAME/KEY: Modified-site  
; LOCATION: 4  
; OTHER INFORMATION: /product= "Acp"  
; FEATURE:  
; NAME/KEY: Modified-site

us-09-394-019a-248.ra

Tue Jul 16 16:10:13 2002

QY 1 KDPXGLEHGDINGXPKGY 18  
||| | |||||  
Db 1 KDXGDEVGDINGXPKGY 18

RESULT 3  
US-08-802-981-69  
; Sequence 69, Application US/08802981  
; Patent No. 6037137  
; GENERAL INFORMATION:  
; APPLICANT: Komoriya, Akira  
; APPLICANT: Packard, Beverly S.  
; TITLE OF INVENTION: Compositions for the Detection of Enzyme  
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 231  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/802,981  
; FILING DATE: 20-FEB-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 016865-0003000US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 69:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 3  
; OTHER INFORMATION: /product= "Aib"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 4  
; OTHER INFORMATION: /product= "Acp"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 14  
; OTHER INFORMATION: /product= "Acp"  
; US-08-802-981-69

Query Match 63.5%; Score 61; DB 3; Length 18;  
Best Local Similarity 77.8%; Pred. No. 0.0003;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
||| | |||||  
Db 1 KDXGDEVGDINGXPKGY 18

RESULT 4  
US-08-802-981-53  
; Sequence 53, Application US/08802981  
; Patent No. 6037137

LOCATION: 14  
OTHER INFORMATION: /product= "Acp"  
US-08-802-981-55

Query Match 67.7%; Score 65; DB 3; Length 18;  
Best Local Similarity 83.3%; Pred. No. 6.2e-05;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
||| | |||||  
Db 1 KDXGDEVGDINGXPKGY 18

RESULT 2  
US-08-802-981-54  
; Sequence 54, Application US/08802981  
; Patent No. 6037137  
; GENERAL INFORMATION:  
; APPLICANT: Komoriya, Akira  
; APPLICANT: Packard, Beverly S.  
; TITLE OF INVENTION: Compositions for the Detection of Enzyme  
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 231  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/802,981  
; FILING DATE: 20-FEB-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 016865-0003000US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 54:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 3  
; OTHER INFORMATION: /product= "Aib"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 4  
; OTHER INFORMATION: /product= "Acp"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 14  
; OTHER INFORMATION: /product= "Acp"  
; US-08-802-981-54

Query Match 66.7%; Score 64; DB 3; Length 18;  
Best Local Similarity 83.3%; Pred. No. 9.2e-05;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

: GENERAL INFORMATION:
: APPLICANT: Komoriya, Akira
: APPLICANT: Packard, Beverly S.
: TITLE OF INVENTION: Compositions for the Detection of Enzyme
: TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
: NUMBER OF SEQUENCES: 231
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/802,981
: FILING DATE: 20-FEB-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Hunter, Tom
: REGISTRATION NUMBER: 38,498
: REFERENCE/DOCKET NUMBER: 016865-000300US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 53:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 18 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: FEATURE:
: NAME/KEY: Modified-site
: LOCATION: 3
: OTHER INFORMATION: /product= "Alb"
: NAME/KEY: Modified-site
: LOCATION: 4
: OTHER INFORMATION: /product= "Acp"
: FEATURE:
: NAME/KEY: Modified-site
: LOCATION: 14
: OTHER INFORMATION: /product= "Acp"
: US-08-802-981-53

Query Match 62.5% Score 60; DB 3; Length 18;
Best Local Similarity 77.8% Pred. No. 0.00044;
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 KDPXGLEHDGNGXPKGY 18
    || || ||| ||| ||| |||
Db .1 KDXXGNEVDGIDGXPKGY 18

RESULT 5
US-08-802-981-57
: Sequence 57, Application US/08802981
: Patent No. 6037137
: GENERAL INFORMATION:
: APPLICANT: Komoriya, Akira
: APPLICANT: Packard, Beverly S.
: TITLE OF INVENTION: Compositions for the Detection of Enzyme
: TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
: NUMBER OF SEQUENCES: 231
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor

```

us-09-394-019a-248.ra1

Tue Jul 16 16:10:13 2002

TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /product= "Aib"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /product= "Acp"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 14  
OTHER INFORMATION: /product= "Acp"  
US-08-802-981-66

Query Match 58.3%; Score 56; DB 3; Length 18;  
Best Local Similarity 72.2%; Pred. No. 0.0021;  
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
Db 1 KDXXGVVADGIDGXPXGY 18

RESULT 8  
US-08-802-981-68  
Sequence 68, Application US/08802981  
Patent No. 6037137  
GENERAL INFORMATION:  
APPLICANT: Komoriya, Akira  
APPLICANT: Packard, Beverly S.  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM: disk  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-0003000US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-0003000US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /product= "Aib"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /product= "Acp"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 14  
OTHER INFORMATION: /product= "Acp"  
US-08-802-981-56

Query Match 61.5%; Score 59; DB 3; Length 18;  
Best Local Similarity 77.8%; Pred. No. 0.00065;  
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
Db 1 KDXXGVVADGIDGXPXGY 18

RESULT 7  
US-08-802-981-66  
Sequence 66, Application US/08802981  
Patent No. 6037137  
GENERAL INFORMATION:  
APPLICANT: Komoriya, Akira  
APPLICANT: Packard, Beverly S.  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM: disk  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-0003000US  
TELECOMMUNICATION INFORMATION:

```
;
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /product= "Aib"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /product= "Acp"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /product= "Acp"
;
; US-08-802-981-68
;
; Query Match      58.38; Score 56; DB 3; Length 18;
; Best Local Similarity 72.28; Pred. No. 0.0021;
; Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
;
; QY 1 KDPXGLEHGDGNGXPKGY 18
;    ||| :|||||
; Db 1 KDXGYVANGINGXPKGY 18
;
; RESULT 9
; US-08-802-981-49
; Sequence 49, Application US/08802981
; Patent No. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; APPLICANT: Packard, Beverly S.
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-000300US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /product= "OTHER"
; OTHER INFORMATION: /note= "Xaa = 9-fluorenylmethoxycarbonyl (Fmoc) group attached to the alpha-amino group of Lys"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
```

```
;
; OTHER INFORMATION: /product= "Aib"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /product= "Acp"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /product= "Acp"
;
; US-08-802-981-49
;
; Query Match      56.28; Score 54; DB 3; Length 18;
; Best Local Similarity 76.58; Pred. No. 0.0046;
; Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
;
; QY 2 DPXGLEHGDGNGXPKGY 18
;    ||| :|||||
; Db 2 DXXGDEVGDGDXPKGY 18
;
; RESULT 10
; US-08-802-981-50
; Sequence 50, Application US/08802981
; Patent No. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; APPLICANT: Packard, Beverly S.
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-000300US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /product= "OTHER"
; OTHER INFORMATION: /note= "Xaa = benzylloxycarbonyl (Z) group attached to the
; OTHER INFORMATION: group of Lys"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /product= "Aib"
; FEATURE:
; NAME/KEY: Modified-site
```

us-09-394-019a-248.ra1

Tue Jul 16 16:10:13 2002

Query Match 53.1%; Score 51; DB 3; Length 18;  
Best Local Similarity 66.7%; Pred. No. 0.015;  
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
||| :||:|||||  
Db 1 KDXXGVVANGIDGXPKGY 18

RESULT 12

US-08-802-981-59  
Sequence 59, Application US/08802981  
Patent No. 6037137

GENERAL INFORMATION:

APPLICANT: Komoriya, Akira

APPLICANT: Packard, Beverly S.

TITLE OF INVENTION: Compositions for the Detection of Enzyme

TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof

NUMBER OF SEQUENCES: 231

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/802,981

FILING DATE: 20-FEB-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hunter, Tom

REGISTRATION NUMBER: 38,498

REFERENCE/DOCKET NUMBER: 016865-00030005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 19 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 3

OTHER INFORMATION: /product= "Aib"

FEATURE:

NAME/KEY: Modified-site

LOCATION: 4

OTHER INFORMATION: /product= "Acp"

FEATURE:

NAME/KEY: Modified-site

LOCATION: 15

OTHER INFORMATION: /product= "Acp"

US-08-802-981-59

Query Match 52.6%; Score 50.5; DB 3; Length 19;

Best Local Similarity 73.7%; Pred. No. 0.019; 3; Indels 1; Gaps 1;

Matches 14; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 KDPXGLEHGDINGXPKGY 18

||| :||:|||||

Db 1 KDXXGVVANGIDGXPKGY 19

;

US-08-802-981-67

Query Match 56.2%; Score 54; DB 3; Length 18;

Best Local Similarity 76.5%; Pred. No. 0.0046; 3; Indels 0; Gaps 0;

Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXGLEHGDINGXPKGY 18

||| :||:|||||

Db 2 DXXGDEVDGIDGXPKGY 18

RESULT 11

US-08-802-981-67

Sequence 67, Application US/08802981

Patent No. 6037137

GENERAL INFORMATION:

APPLICANT: Komoriya, Akira

APPLICANT: Packard, Beverly S.

TITLE OF INVENTION: Compositions for the Detection of Enzyme

TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof

NUMBER OF SEQUENCES: 231

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/802,981

FILING DATE: 20-FEB-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hunter, Tom

REGISTRATION NUMBER: 38,498

REFERENCE/DOCKET NUMBER: 016865-00030005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 67:

SEQUENCE CHARACTERISTICS:

LENGTH: 18 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 3

OTHER INFORMATION: /product= "Aib"

FEATURE:

NAME/KEY: Modified-site

LOCATION: 4

OTHER INFORMATION: /product= "Acp"

FEATURE:

NAME/KEY: Modified-site

LOCATION: 14

OTHER INFORMATION: /product= "Acp"



us-09-394-019a-248.rai

Tue Jul 16 16:10:13 2002

```

Best Local Similarity 70.6%; Pred. No. 0.047;
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 DPXGLEHGDINGXPKGY 18
   | | | | | | | | | |
Db 2 DXGDEVGDIDGXPXGY 18

RESULT 15
US-08-802-981-76
; Sequence 76, Application US/08802981
; Patent NO. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; APPLICANT: Packard, Beverly S.
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-0003000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /product= "Aib"
;
US-08-802-981-76

```

```

Query Match 49.5%; Score 47.5; DB 3; Length 19;
Best Local Similarity 52.6%; Pred. No. 0.061;
Matches 10; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

Qy 1 KDPXGLEHD-GINGXPKGY 18
   | | | | | | | | | |
Db 1 KDXGGIETDSGVDDGPKGY 19

```

Search completed: July 16, 2002, 11:20:29  
Job time: 307 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2002, 11:15:22 ; Search time 14.62 Seconds  
(without alignments)  
118.304 Million cell updates/sec

Title: US-09-394-019A-248

Perfect score: 96

Sequence: 1 KDPXGLEHGDINGXPKGY 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: PIR\_71.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	45	46.9	135	2 G87653	hypothetical prote
2	44.5	46.4	542	2 E72212	formate--tetrahydr
3	43.5	45.3	1146	2 T09112	probable sensor ki
4	43.5	45.3	1146	2 A82174	sensory box sensor
5	43	44.8	289	2 A43562	homeotic protein H
6	43	44.8	333	2 S42424	hypothetical prote
7	42.5	44.3	370	2 D83818	heat-shock protein
8	42	43.8	73	2 I47089	keratin type II -
9	42	43.8	99	2 A41605	homeotic protein H
10	42	43.8	118	2 JC2568	myr protein - Rhl
11	42	43.8	129	2 I56195	gene Tap-1 protein
12	42	43.8	196	2 AH2441	hypothetical prote
13	42	43.8	200	2 H82320	HAMI protein VC045
14	42	43.8	427	2 I51580	XFRH2 protein - Af
15	42	43.8	497	1 WMBELM	membrane protein L
16	42	43.8	577	2 A37779	histocompatibility
17	42	43.8	725	2 S13426	multidrug resistan
18	41.5	43.2	327	2 AI0522	pdxA-like protein
19	41	42.7	94	2 F82636	hypothetical prote
20	41	42.7	327	2 G83921	hypothetical prote
21	41	42.7	448	2 A36311	70K UI small nucle
22	41	42.7	519	2 S78850	hypothetical prote
23	41	42.7	798	2 S29815	N-ras upstream pro
24	41	42.7	798	2 S11210	probable unr prote
25	41	42.7	902	2 A55543	cmaA protein - Pse
26	41	42.7	1146	2 B35962	protein-tyrosine k
27	41	42.7	1182	2 A35962	protein-tyrosine k
28	40.5	42.2	1451	2 D64203	DNA polymerase III
29	40	41.7	185	2 T12772	conserved hypothet

## ALIGNMENTS

## RESULT 1

G87653

hypothetical protein CC3265 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: G87653

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, N.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: G87653

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-135 <STO>

A:Cross-references: GB:AE005673; NID:gl3424955; PIDN:AAK25227.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC3265

Query Match 46.9%; Score 45; DB 2; Length 135;  
Best Local Similarity 50.0%; Pred. No. 2.4;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 GLEHGDINGXPKGY 18

I::|||I|||

Db 89 GMDFDGTGKLPGKW 102

## RESULT 2

E72212

formate--tetrahydrofolate ligase - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: E72212

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316

A:Accession: E72212

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-542 <ARN>

A:Cross-references: GB:AE001815; GB:AE000512; NID:g4982341; PIDN:AAD36830.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1766

C:Superfamily: formate--tetrahydrofolate ligase; formate--tetrahydrofolate ligase hom

```

A43562
homeotic protein Hox D8 - mouse
N;Alternate names: homeotic protein Hox 4.3
C;Species: Mus musculus (house mouse)
C;Date: 11-Dec-1992 #sequence_revision 11-Dec-1992 #text_change 30-May-1997
C;Accession: A43562
R;Izpisua-Belmonte, J.C.; Dolle, P.; Renucci, A.; Zappavigna, V.; Falkenstein, H.; Du
Development 110, 733-745, 1990
A;Title: Primary structure and embryonic expression pattern of the mouse Hox-4.3 home
A;Reference number: A43562; MUID:91209232
A;Accession: A43562
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-289 <IZP>
A;Superfamily: homeotic protein Hox A7; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;196-252/Domain: homeobox homology <HOX>

Query Match 44.8%; Score 43; DB 2; Length 289;
Best Local Similarity 53.3%; Pred. No. 12;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 KDPXGLEHDXGPKY 15
      ||| ||| |||
DB 273 KEASGLEEDGAGGCP 287

RESULT 6
S42424
hypothetical protein Y - Mycobacterium smegmatis
C;Species: Mycobacterium smegmatis
C;Date: 07-May-1998 #sequence_revision 15-May-1998 #text_change 22-Oct-1999
C;Accession: S42424; S31804
R;Cirillo, J.D.; Weisbrod, T.R.; Pascopella, L.; Bloom, B.R.; Jacobs Jr., W.R.
Mol. Microbiol. 11, 629-639, 1994
A;Title: Isolation and characterization of the aspartokinase and aspartate semialdehy
A;Reference number: S42421; MUID:94254720
A;Accession: S42424
A;Molecule type: DNA
A;Residues: 1-333 <CIR>
A;Cross-references: EMBL:Z17372; NID:644506; PIDN:CAA78987.1; PID:g581353
A;Note: the authors translated the initiation codon GTG for residue 1 as Val
C;Genetics:
A;Start codon: GTG

Query Match 44.8%; Score 43; DB 2; Length 333;
Best Local Similarity 40.0%; Pred. No. 15;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 DPXGLEHDXGPKY 16
      ||| ||| ||| |||
DB 67 DPAGIRYGVGVGDK 81

RESULT 7
D83818
heat-shock protein (activation of DnaK) dnaJ [imported] - Bacillus halodurans (strain
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: D83818
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: D83818
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-370 <STO>
A;Cross-references: GB:AP001511; GB:BA000004; NID:g1017372; PIDN:BA05067.1; GSPDB:G
C;Experimental source: strain C-125
C;Genetics:

```

```

A43562
homeotic protein Hox D8 - mouse
N;Alternate names: homeotic protein Hox 4.3
C;Species: Mus musculus (house mouse)
C;Date: 11-Dec-1992 #sequence_revision 11-Dec-1992 #text_change 30-May-1997
C;Accession: A43562
R;Izpisua-Belmonte, J.C.; Dolle, P.; Renucci, A.; Zappavigna, V.; Falkenstein, H.; Du
Development 110, 733-745, 1990
A;Title: Primary structure and embryonic expression pattern of the mouse Hox-4.3 home
A;Reference number: A43562; MUID:91209232
A;Accession: A43562
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-289 <IZP>
A;Superfamily: homeotic protein Hox A7; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;196-252/Domain: homeobox homology <HOX>

Query Match 46.4%; Score 44.5; DB 2; Length 542;
Best Local Similarity 42.1%; Pred. No. 14;
Matches 8; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 1 KDPXGLEHD-GINGXPKGY 18
      ||| ||| ||| ||| |||
DB 474 KTKPSISHDPSLRGAPGY 492

RESULT 3
T09112
probable sensor kinase Vies - Vibrio cholerae
C;Species: Vibrio cholerae
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: T09112
R;Lee, S.H.; Angelichio, M.J.; Mekalanos, J.J.; Camilli, A.
J. Bacteriol. 180, 2298-2305, 1998
A;Title: Nucleotide sequence and spatio-temporal expression studies of the Vibrio cholera
A;Reference number: Z16568; MUID:98233722
A;Accession: T09112
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1146 <LEE>
A;Cross-references: EMBL:AF031552; NID:g3123888; PIDN:AAC38448.1; PID:g3123890
A;Experimental source: strain C6709-1; El Tor
C;Genetics:
A;Gene: vies
C;Keywords: signal transduction

Query Match 45.3%; Score 43.5; DB 2; Length 1146;
Best Local Similarity 37.0%; Pred. No. 50;
Matches 10; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

QY 1 KDPXGLEHD-----GINGXPKGY 18
      ||| ||| ||| |||
DB 9 QEPGIEHDFASGIAKELGINIEYKGF 35

RESULT 4
A82174
sensory box sensor histidine kinase/response regulator Vies VC1653 [imported] - Vibrio c
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: A82174
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: A82174
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1146 <HEI>
A;Cross-references: GB:AE004243; GB:AE003852; NID:g9656163; PIDN:AAF94804.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC1653
A;Map position: 1

Query Match 45.3%; Score 43.5; DB 2; Length 1146;
Best Local Similarity 37.0%; Pred. No. 50;
Matches 10; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

QY 1 KDPXGLEHD-----GINGXPKGY 18
      ||| ||| ||| |||
DB 9 QEPGIEHDFASGIAKELGINIEYKGF 35

RESULT 5

```

A:Gene: dnaJ  
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

Query Match 44.3%; Score 42.5; DB 2; Length 370;  
Best Local Similarity 32.1%; Pred. No. 20;  
Matches 9; Conservative 3; Mismatches 5; Indels 11; Gaps 1;

QY 1 KDPXGLEHD-----GKXPKG 17  
| | | | | : | | | | |  
Db 216 KVPAGIDHGQIRLSGGGAGVNGGPAG 243

RESULT 8  
I47089  
keratin type II - sheep (fragment)  
C:Species: Ovis sp. (sheep)  
C:Date: 15-Oct-1996 #sequence\_revision 15-Oct-1996 #text\_change 13-Aug-1999  
C:Accession: I47089  
R:Powell, B.C.; Cam, G.R.; Fietz, M.J.; Rogers, G.E.  
Proc. Natl. Acad. Sci. U.S.A. 83, 5048-5052, 1986  
A:Title: Clustered arrangement of keratin intermediate filament genes.  
A:Reference number: I47089; MUID:86259729  
A:Accession: I47089  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-73 <POW>  
A:Cross-references: GB:M13679; NID:g165954; PIDN:AAA31555.1; PID:g552427  
C:Superfamily: cytoskeletal keratin

Query Match 43.8%; Score 42; DB 2; Length 73;  
Best Local Similarity 41.2%; Pred. No. 3.8;  
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 KDPXGLEHDGKXPKG 17  
| | | | | : | | | | |  
Db 8 KDPALQESASGSAK 24

RESULT 9  
A41605  
homeotic protein Hox 4.3 - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Jun-1992 #sequence\_revision 01-Dec-1995 #text\_change 24-Sep-1999  
C:Accession: S16177; C42694; A41605; A38810; S15521  
R:Sadoul, R.; Featherstone, M.S.  
Biochim. Biophys. Acta 1089, 259-261, 1991  
A:Title: Sequence analysis of the homeobox-containing exon of the murine Hox-4.3 homeo-  
A:Reference number: S16177; MUID:91274361  
A:Accession: S16177  
A:Molecule type: DNA  
A:Residues: 1-99 <BIO>  
A:Cross-references: EMBL:X56561; NID:g51418; PIDN:CAA39911.1; PID:e30697; PID:g1333935  
R:Nazarali, A.; Kim, Y.; Nirenberg, M.  
Proc. Natl. Acad. Sci. U.S.A. 89, 2883-2887, 1992  
A:Title: Hox-1.1 and Hox-4.9 homeobox genes.  
A:Reference number: A42694; MUID:92212934  
A:Accession: C42694  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 2-70 <NAZ>  
A:Cross-references: GB:M87803; NID:g193953; PIDN:AAA37852.1; PID:g193954  
R:Singh, G.; Kaur, S.; Stock, J.L.; Jenkins, N.A.; Gilbert, D.J.; Copeland, N.G.; Potter  
Proc. Natl. Acad. Sci. U.S.A. 88, 10706-10710, 1991  
A:Title: Identification of 10 murine homeobox genes.  
A:Reference number: A37290; MUID:92073356  
A:Accession: A41605  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 5-64 <SI2>  
C:Genetics:

A:Gene: Hox 4.3  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:6-62/Domain: homeobox homology <HOX>

Query Match 43.8%; Score 42; DB 2; Length 99;  
Best Local Similarity 53.3%; Pred. No. 5.4;  
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1' KDPXGLEHDGKXPKG 15  
| | | | | : | | | | |  
Db 83 KEVSGLEDGAGCPC 97

RESULT 10  
JC2568  
mray protein - Rhizobium meliloti (fragment)  
C:Species: Rhizobium meliloti  
C:Date: 13-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 03-Nov-2000  
C:Accession: JC2568  
R:Leach, F.; Wacker, D.B.; Signer, E.R.  
Gene 148, 87-90, 1994  
A:Title: Rhizobium meliloti homologs of Escherichia coli mur genes.  
A:Reference number: JC2567; MUID:95011665  
A:Accession: JC2568  
A:Molecule type: DNA  
A:Residues: 1-118 <LEA>  
A:Cross-references: GB:L25875  
C:Genetics:  
A:Gene: mray  
C:Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase

Query Match 43.8%; Score 42; DB 2; Length 118;  
Best Local Similarity 43.8%; Pred. No. 6.7;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 DPXGLEHDGKXPKG 17  
| | | | | : | | | | |  
Db 96 DPPSLREGLDGKPGG 111

RESULT 11  
I56195  
gene Tap-1 protein - mouse (fragment)  
C:Species: Mus sp. (mouse)  
C:Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 20-Sep-1999  
C:Accession: I56195  
R:Pearce, R.B.; Triglier, L.; Svaasand, E.K.; Peterson, C.M.  
J. Immunol. 151, 5338-5347, 1993  
A:Title: Polymorphism in the mouse Tap-1 gene. Association with abnormal CD8+ T cell  
A:Reference number: I56195; MUID:94044758  
A:Accession: I56195  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-129 <RES>  
A:Cross-references: GB:S66882; NID:g440949; PIDN:AAB28745.1; PID:g440950  
C:Genetics:  
A:Gene: Tap-1  
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog  
C:Keywords: ATP  
F:1-129/Domain: ATP-binding cassette homology (fragment) <ABC>

Query Match 43.8%; Score 42; DB 2; Length 129;  
Best Local Similarity 63.6%; Pred. No. 7.4;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 8 HDGKXPKGY 18  
| | | | | : | | | | |  
Db 79 HDFISGFPOGY 89

us-09-394-019a-248.rpr

Tue Jul 16 16:10:14 2002

Dev. Biol. 160, 413-423, 1993  
 A:Title: XFKH2, a Xenopus HNF-3 alpha homologue, exhibits both activin-inducible and  
 A:Reference number: I51580; MUID:94074768  
 A:Accession: I51580  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-427 <BOL>  
 A:Cross-references: GB:M93658; NID:g214901; PIDN:AAA17050.1; PID:g214902  
 C:Genetics:  
 A:Gene: XFKH2  
 C:Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology  
 F:157-248/Domain: fork head DNA-binding domain homology <FHD>

Query Match 43.8%; Score 42; DB 2; Length 427;  
 Best Local Similarity 57.1%; Pred. No. 29;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 PXGLEHGDGNGXPK 16  
 | | | | | | | | | |  
 Db 301 PQALEHNGSGEMK 314

## RESULT 15

WMBELM

membrane protein LMP-2A - human herpesvirus 4  
 N:Contains: membrane protein LMP-2B  
 C:Species: human herpesvirus 4, Epstein-Barr virus  
 A:Note: host Homo sapiens (man)  
 C:Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 16-Jul-1999  
 C:Accession: A30178; B30178; S00392  
 R:Sample, J.; Liebowitz, D.; Kieff, E.  
 J. Virol. 63, 933-937, 1989  
 A:Title: Two related Epstein-Barr virus membrane proteins are encoded by separate gen  
 A:Reference number: A30178; MUID:89095024  
 A:Accession: A30178  
 A:Molecule type: mRNA  
 A:Residues: 1-497 <SMA>  
 A:Cross-references: GB:M24212; NID:g522186; PIDN:AAA45887.1; PID:g522187  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 120-497 <SA2>  
 R:Laux, G.; Perricaudet, M.; Farrell, P.J.  
 EMBO J. 7, 769-774, 1988

QY 2 DPXGLEHGDGNGXPKGY 18  
 | | | | | | | | | |  
 Db 67 DDSGLQVDALNGLPGVY 83

RESULT 13  
 HAM1 protein VC0456 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
 C:Species: Vibrio cholerae  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A:Reference number: A82035; MUID:20406833  
 A:Accession: H82320  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-200 <HEI>  
 A:Cross-references: GB:AE004132; GB:AE003852; NID:g9654871; PIDN:AAF93629.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Map position: 1  
 A:Gene: VC0456  
 C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0226

Query Match 43.8%; Score 42; DB 2; Length 200;  
 Best Local Similarity 52.9%; Pred. No. 12;  
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 DPXGLEHGDGNGXPKGY 18  
 | | | | | | | | | |  
 Db 68 DDSGLQVDALNGLPGVY 84

## RESULT 14

I51580

XFKH2 protein - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 16-Jul-1999  
 C:Accession: I51580  
 R:Polce, M.E.; Hemmati-Brivanlou, A.; Harland, R.M.

Query Match

43.8%; Score 42; DB 1; Length 497;

Tue Jul 16 16:10:14 2002

us-09-394-019a-248.rpr

Best Local Similarity 63.6%; Pred. No. 34;  
Matches 7; Conservative 2; Mismatches

2; Indels 0; Gaps 0;

QY 5 GLEHGDGNGXP 15

|||||

Db 87 GLQHDGNDGLP 97

Search completed: July 16, 2002, 11:20:52  
Job time: 330 sec

**THIS PAGE BLANK (copy)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2002, 11:20:22 ; Search time 10.36 Seconds  
(without alignments)  
67.273 Million cell updates/sec

Title: US-09-394-019A-248

Perfect score: 96

Sequence: 1 KDPXGLEHNGXPKGY 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43.5	45.3	655	1 ABG2_HUMAN	O9unq0 homo sapien
2	43	44.8	289	1 HXD8_MOUSE	P23463 mus musculus
3	43	44.8	333	1 YASD_MYCSM	P41402 mycobacteri
4	43	44.8	352	1 NDR4_HUMAN	Q9ulp0 homo sapien
5	43	44.8	352	1 NDR4_RAT	Q92219 rattus norv
6	42.5	44.3	370	1 DNAJ_BACHD	O9kd71 bacillus ha
7	42	43.8	282	1 YC80_GUITH	O78449 guillardia
8	42	43.8	375	1 NDR3_MOUSE	Q9qvf9 mus musculus
9	42	43.8	427	1 FKH2_XENLA	P32315 xenopus lae
10	42	43.8	497	1 LMP2_EBV	P13285 Epstein-bar
11	42	43.8	724	1 TAP1_MOUSE	P21958 mus musculus
12	42	43.8	725	1 TAP1_RAT	P36370 rattus norv
13	41.5	43.2	327	1 PXA2_SALTI	P58716 salmonella
14	41.5	43.2	327	1 PXA2_SALTY	P58718 salmonella
15	41	42.7	448	1 RU17_DROME	P17133 drosophila
16	41	42.7	798	1 UNR_HUMAN	O75534 homo sapien
17	41	42.7	798	1 UNR_RAT	P18395 rattus norv
18	41	42.7	1182	1 ABL2_HUMAN	P42684 homo sapien
19	40.5	42.2	1451	1 DPO3_MYCGE	P47277 mycoplasma
20	40	41.7	375	1 PKG_TETPY	O00871 tetrahymena
21	40	41.7	481	1 BIND_STRPU	P06651 strongyloce
22	39.5	41.1	397	1 YEDS_ECOLI	P76335 escherichia
23	39	40.6	289	1 DLX5_MOUSE	P70396 mus musculus
24	39	40.6	471	1 DNAB_SALTY	P10338 salmonella
25	39	40.6	583	1 MLIX_MOUSE	O88495 mus musculus
26	39	40.6	787	1 SYFB_NEIMA	Q9jva0 neisseria m
27	39	40.6	787	1 SYFB_NEIMB	Q9k089 neisseria m
28	38	39.6	309	1 Y922_RHILO	Q98h78 rhizobium l
29	38	39.6	379	1 Y922_HAEIN	P44544 haemophilus
30	38	39.6	459	1 DNAB_BUCAI	P57611 buchonia ap
31	38	39.6	471	1 DNAB_ECOLI	P03005 escherichia
32	38	39.6	525	1 HRG_HUMAN	P04196 homo sapien
33	38	39.6	574	1 KPR_RAT	P12928 rattus norv

34	38	39.6	1763	1 POLN_FCVF9	P27409 feline call
35	37.5	39.1	650	1 ABG3_MOUSE	O99p81 mus musculu
36	37.5	39.1	922	1 PMPI_CHLPN	Q92945 chlamydia p
37	37	38.5	136	1 GLB3_CHITP	P22431 chironomus
38	37	38.5	151	1 GLB3_CHITH	P02229 chironomus
39	37	38.5	151	1 GLB4_CHITH	P02230 chironomus
40	37	38.5	222	1 YP32_YEAST	P02230 chironomus
41	37	38.5	248	1 HMD1_CHICK	P51996 saccharomyc
42	37	38.5	302	1 LGUL_SCHPO	P46692 gallus gall
43	37	38.5	334	1 CHMU_ARATH	Q09751 schizosacch
44	37	38.5	352	1 NIVO_CLOPA	P42738 arabidopsis
45	37	38.5	356	1 MPPC_RAT	Q00852 clostridium
					P16036 rattus norv

## ALIGNMENTS

RESULT 1  
ABG2\_HUMAN  
ID ABG2\_HUMAN STANDARD; PRT; 655 AA.  
AC Q9UNQ0; O95374; O9NUS0; O9BY73;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-binding cassette transporter) (Breast cancer resistance protein).  
GN ABCG2 OR ABCP OR BCRP OR BCRP1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=99065313; PubMed=9850061;  
RA Allikmets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;  
RT "A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance.";  
RL Cancer Res. 58:5337-5339(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast cancer;  
RX MEDLINE=99080071; PubMed=9861027;  
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,  
Ross D.D.;  
RT "A multidrug resistance transporter from human MCF-7 breast cancer cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).  
RN [3]  
RP ERRATUM.  
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,  
Ross D.D.;  
RT "Breast cancer resistance protein constitutes a 140-kDa complex as a homodimer.";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 198-655 FROM N.A.  
RC TISSUE=Placenta;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
Wagatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
Takanashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,  
Ninomiya K., Iwayanagi T.;  
RT "NED0 human cDNA sequencing project.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]

01-NOV-1997 (Rel. 35, Last annotation update)  
Homeobox protein Hox-D8 (Hox-4.3) (Hox-5.4).  
HOXD8 OR HOXD-8 OR Hox-4.3.  
Mus musculus (Mouse).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_Taxid=10090;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=91209232; PubMed=1982431;  
Izpisua-Belmonte J.-C., Dolle P., Renucci A., Zappavigna V.,  
Falkenstein H., Duboule D.;  
Primary structure and embryonic expression pattern of the mouse  
Hox-4.3 homeobox gene.;  
Development 110:733-745(1990).  
[2]  
SOURCE OF 191-289 FROM N.A.  
MEDLINE=91274361; PubMed=1675873;  
Sedoul R., Featherstone M.;  
Sequence analysis of the homeobox-containing exon of the murine  
Hox-4.3 homeogene.;  
Biochim. Biophys. Acta 1089:259-261(1991).  
[3]  
SEQUENCE OF 195-254 FROM N.A.  
MEDLINE=92073356; PubMed=1683707;  
Singh G., Kaur S., Stock J.L., Jenkins N.A., Gilbert D.J.,  
Copeland N.G., Potter S.S.;  
Identification of 10 murine homeobox genes.;  
Proc. Natl. Acad. Sci. U.S.A. 88:10706-10710(1991).  
[4]  
SEQUENCE OF 192-260 FROM N.A.  
MEDLINE=92212934; PubMed=1348361;  
Nazari A., Kim Y., Nirenberg M.;  
Hox-1.11 and Hox-4.9 homeobox genes.;  
Proc. Natl. Acad. Sci. U.S.A. 89:2883-2887(1992).  
!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
!- SUBCELLULAR LOCATION: Nuclear.  
!- SIMILARITY: BELONGS TO THE ANT-P FAMILY OF HOMEBOX PROTEINS.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).

EMBL; X56561; CAA39911.1; -  
EMBL; M87803; AAA37852.1; -  
PIR; A38810; A38810.  
PIR; A41605; A41605.  
PIR; A43562; A43562.  
PIR; S16177; S16177.  
PIR; S15521; S15521.  
DR HSP; P02833; 9ANT.  
TRANSFAC; T01426; -  
MGD; MG1:96209; Hoxd8.  
InterPro; IPR001827; Antennapedia.  
DR InterPro; IPR001356; Homeobox.  
PFam; PF00046; homeobox; 1.  
PRINTS; PR00025; ANTENNAPEDIA.  
DR PRINTS; PR00024; HOMEBOX.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00032; ANTENNAPEDIA; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
Transcription regulation.  
DOMAIN 15 28 POLY-ALA.  
FT DOMAIN 62 89 GLY/PRO-RICH.  
FT DOMAIN 108 117 POLY-PRO.

DT  
DE Homeobox protein Hox-D8 (Hox-4.3) (Hox-5.4).  
GN HOXD8 OR HOXD-8 OR Hox-4.3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91209232; PubMed=1982431;  
RA Izpisua-Belmonte J.-C., Dolle P., Renucci A., Zappavigna V.,  
FA Falkenstein H., Duboule D.;  
RT Primary structure and embryonic expression pattern of the mouse  
Hox-4.3 homeobox gene.;  
RL Development 110:733-745(1990).  
RN [2]  
RP SOURCE OF 191-289 FROM N.A.  
RX MEDLINE=91274361; PubMed=1675873;  
RA Sedoul R., Featherstone M.;  
RT Sequence analysis of the homeobox-containing exon of the murine  
Hox-4.3 homeogene.;  
RL Biochim. Biophys. Acta 1089:259-261(1991).  
RN [3]  
RP SEQUENCE OF 195-254 FROM N.A.  
RX MEDLINE=92073356; PubMed=1683707;  
RA Singh G., Kaur S., Stock J.L., Jenkins N.A., Gilbert D.J.,  
CO Copeland N.G., Potter S.S.;  
RT Identification of 10 murine homeobox genes.;  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10706-10710(1991).  
RN [4]  
RP SEQUENCE OF 192-260 FROM N.A.  
RX MEDLINE=92212934; PubMed=1348361;  
RA Nazari A., Kim Y., Nirenberg M.;  
RT Hox-1.11 and Hox-4.9 homeobox genes.;  
RL Proc. Natl. Acad. Sci. U.S.A. 89:2883-2887(1992).  
CC !- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
CC !- SUBCELLULAR LOCATION: Nuclear.  
CC !- SIMILARITY: BELONGS TO THE ANT-P FAMILY OF HOMEBOX PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).

EMBL; X56561; CAA39911.1; -  
EMBL; M87803; AAA37852.1; -  
PIR; A38810; A38810.  
PIR; A41605; A41605.  
PIR; A43562; A43562.  
PIR; S16177; S16177.  
PIR; S15521; S15521.  
DR HSP; P02833; 9ANT.  
TRANSFAC; T01426; -  
MGD; MG1:96209; Hoxd8.  
InterPro; IPR001827; Antennapedia.  
DR InterPro; IPR001356; Homeobox.  
PFam; PF00046; homeobox; 1.  
PRINTS; PR00025; ANTENNAPEDIA.  
DR PRINTS; PR00024; HOMEBOX.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00032; ANTENNAPEDIA; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
Transcription regulation.  
DOMAIN 15 28 POLY-ALA.  
FT DOMAIN 62 89 GLY/PRO-RICH.  
FT DOMAIN 108 117 POLY-PRO.

DT  
DE Homeobox protein Hox-D8 (Hox-4.3) (Hox-5.4).  
GN HOXD8 OR HOXD-8 OR Hox-4.3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91209232; PubMed=1982431;  
RA Izpisua-Belmonte J.-C., Dolle P., Renucci A., Zappavigna V.,  
FA Falkenstein H., Duboule D.;  
RT Primary structure and embryonic expression pattern of the mouse  
Hox-4.3 homeobox gene.;  
RL Development 110:733-745(1990).  
RN [2]  
RP SOURCE OF 191-289 FROM N.A.  
RX MEDLINE=91274361; PubMed=1675873;  
RA Sedoul R., Featherstone M.;  
RT Sequence analysis of the homeobox-containing exon of the murine  
Hox-4.3 homeogene.;  
RL Biochim. Biophys. Acta 1089:259-261(1991).  
RN [3]  
RP SEQUENCE OF 195-254 FROM N.A.  
RX MEDLINE=92073356; PubMed=1683707;  
RA Singh G., Kaur S., Stock J.L., Jenkins N.A., Gilbert D.J.,  
CO Copeland N.G., Potter S.S.;  
RT Identification of 10 murine homeobox genes.;  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10706-10710(1991).  
RN [4]  
RP SEQUENCE OF 192-260 FROM N.A.  
RX MEDLINE=92212934; PubMed=1348361;  
RA Nazari A., Kim Y., Nirenberg M.;  
RT Hox-1.11 and Hox-4.9 homeobox genes.;  
RL Proc. Natl. Acad. Sci. U.S.A. 89:2883-2887(1992).  
CC !- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
CC !- SUBCELLULAR LOCATION: Nuclear.  
CC !- SIMILARITY: BELONGS TO THE ANT-P FAMILY OF HOMEBOX PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).

EMBL; X56561; CAA39911.1; -  
EMBL; M87803; AAA37852.1; -  
PIR; A38810; A38810.  
PIR; A41605; A41605.  
PIR; A43562; A43562.  
PIR; S16177; S16177.  
PIR; S15521; S15521.  
DR HSP; P02833; 9ANT.  
TRANSFAC; T01426; -  
MGD; MG1:96209; Hoxd8.  
InterPro; IPR001827; Antennapedia.  
DR InterPro; IPR001356; Homeobox.  
PFam; PF00046; homeobox; 1.  
PRINTS; PR00025; ANTENNAPEDIA.  
DR PRINTS; PR00024; HOMEBOX.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00032; ANTENNAPEDIA; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
Transcription regulation.  
DOMAIN 15 28 POLY-ALA.  
FT DOMAIN 62 89 GLY/PRO-RICH.  
FT DOMAIN 108 117 POLY-PRO.

DT  
DE Homeobox protein Hox-D8 (Hox-4.3) (Hox-5.4).  
GN HOXD8 OR HOXD-8 OR Hox-4.3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91209232; PubMed=1982431;  
RA Izpisua-Belmonte J.-C., Dolle P., Renucci A., Zappavigna V.,  
FA Falkenstein H., Duboule D.;  
RT Primary structure and embryonic expression pattern of the mouse  
Hox-4.3 homeobox gene.;  
RL Development 1



```
FT DOMAIN 183 188 ANTP-TYPE HEXAPEPTIDE.
FT DNA_BIND 195 254 HOMEOBOX.
FT CONFLICT 207 208 TL -> RV (IN REF. 1).
FT CONFLICT 231 231 T -> S (IN REF. 1).
FT CONFLICT 265 266 EA -> DG (IN REF. 1).
FT CONFLICT 275 275 A -> V (IN REF. 2).
SQ SEQUENCE 289 AA; 31410 MW; 5783099FB9B2BDFE CRC64;

Query Match 44.88; Score 43; DB 1; Length 289;
Best Local Similarity 53.3%; Pred. No. 6.8;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 KDPXGLEHDGNGXP 15
   I: I I I I I
Db 273 KEASGLEEDGAECCP 287

RESULT 3
YASD_MYCSM
ID YASD_MYCSM STANDARD; PRT; 333 AA.
AC P41402;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 35.9 kDa protein in ASD 3'region (ORFV).
OS Mycobacterium smegmatis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID:1772;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 607 / MC(2)6 / NRRL B-692;
RX MEDLINE=94254720; PubMed=7910936;
RA Cirillo J.D., Weisbrod T.R., Pascopella L., Bloom B.R.,
RA Jacobs W.R., Jr.;
RT "Isolation and characterization of the aspartokinase and aspartate
RT semialdehyde dehydrogenase operon from mycobacteria.";
RL Mol. Microbiol. 11:629-639(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z17372; CAA78987.1; -
KW Hypothetical protein.
SQ SEQUENCE 333 AA; 35881 MW; C1C50D9A6DC37368 CRC64;

Query Match 44.88; Score 43; DB 1; Length 333;
Best Local Similarity 40.0%; Pred. No. 8;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 DPXGLEHDGNGXPK 16
   I: I: I: I: I
Db 67 DPAGIRYNGVSGVDK 81

RESULT 4
NDRA_HUMAN
ID NDRA4_HUMAN STANDARD; PRT; 352 AA.
AC Q9ULP0; Q9GZX0; Q9GZN3; Q9GZM1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE NDRA4 protein (Brain development-related molecule 1) (Vascular smooth
DE muscle cell associated protein-8) (SMAP-8).
GN NDRA4 OR BDW1 OR KIAA1180.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
RA Zhou R.-H., Kokame K., Tsukamoto Y., Yutani C., Kato H., Miyata T.;
RT "Characterization of the human NDRG gene family: a newly identified
RT member, NDRG4, is specifically expressed in brain and heart.";
RL Genomics 73:86-97(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Heart;
RA Nishimoto S., Tawara J., Aoki T., Toyoda H., Komuraaki T.;
RT "Molecular cloning and characterization of the human vascular smooth
RT muscle cell associated protein-8 (SMAP-8).";
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Amygdala;
RA Wiemann S., Weill B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansoerge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Meves H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wanbutt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
RN [4]
RP SEQUENCE OF 80-352 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20039618; PubMed=10574461;
RA Hiroseawa M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,
RA Ohara O.;
RT "Characterization of cDNA clones selected by the GeneMark analysis
RT from size-fractionated cDNA libraries from human brain.";
RL DNA Res. 6:329-336(1999).
CC -!- FUNCTION: PLAY A ROLE IN THE EARLY POSTNATAL DEVELOPMENT AND
CC -!- FUNCTION OF NEURONAL CELLS (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS: 3 isoforms; 1/NDRG4-Bvar (shown here),
CC 2/NDRG4-B and 3/NDRG4-H; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Expressed only in brain and heart. Isoforms 1
CC and 2 are only expressed in brain. Isoform 3 is expressed in both
CC heart and brain.
CC -!- SIMILARITY: BELONGS TO THE NDRG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB044947; BAB20071.1; -
DR EMBL: AB044947; BAB20072.1; -
DR EMBL: AB044947; BAB20073.1; -
DR EMBL: AB044944; BAB20068.1; -
DR EMBL: AB044945; BAB20069.1; -
DR EMBL: AB044946; BAB20070.1; -
DR EMBL: AB021172; BAB20288.1; -
DR EMBL: AL136584; CAB66519.1; -
DR EMBL: AB033006; BAA86494.1; -
DR InterPro: IPR004142; Ndr.
DR Pfam: PF03096; Ndr. 1.
KW Developmental protein; Alternative splicing.
FT DOMAIN 253 256 POLY-THR.
FT VARSPLIC 1 7 MPECWDG -> MAGLOELRPPEKPLLRQDQATELESSDAF
FT VARSPLIC 289 302 LLAADTDWK (IN ISOFORM 3).
FT VARSPLIC 302 IAYLKRRLRSGGAV -> M (IN ISOFORM 2 AND
FT ISOFORM 3).
SQ SEQUENCE 352 AA; 38459 MW; 4CFAFC7C820013B6 CRC64;
```

Tue Jul 16 16:10:14 2002

DT	16-OCT-2001 (Rel. 40, Last annotation update)	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Chaperone protein dnaJ	
GN	DNAJ OR BH1348	
OS	Bacillus halodurans	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/Staphylococcus group; Bacillus	
OX	NCBI_TaxID=86665;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C-125 / JCM 9153;	
RX	MEDLINE=20512582; PubMed=11058132;	
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,	
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,	
RA	Horikoshi K.;	
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus	
RT	halodurans and genomic sequence comparison with Bacillus subtilis."	
RL	Nucleic Acids Res. 28:4317-4331(2000).	
CC	-!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,	
CC	THE ATPASE ACTIVITY OF DNAJ (BY SIMILARITY).	
CC	-!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).	
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).	
CC	-!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	EMBL; AP001511; BAB05067.1;	
DR	InterPro: IPR003095; DnaJ_C.	
DR	InterPro: IPR002939; DnaJ_C.	
DR	InterPro: IPR001305; DnaJ_CXXCXXG.	
DR	InterPro: IPR001623; DnaJ_N.	
DR	Pfam: PF01556; DnaJ_C.1.	
DR	Pfam: PF00684; DnaJ_CXXCXXG.1.	
DR	PRINTS; PR00625; DNAJPROTEIN.	
DR	SMART; SM00271; DnaJ.1.	
DR	PROSITE; PS00636; DnaJ_1; 1.	
DR	PROSITE; PS00076; DnaJ_2; 1.	
DR	PROSITE; PS00637; DnaJ_CXXCXXG.1.	
KW	Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding;	
KW	Complete proteome.	
FT	DOMAIN 5 69	J-DOMAIN.
FT	DOMAIN 72 97	GLY-RICH.
FT	REPEAT 141 148	CXXCXXG MOTIF.
FT	REPEAT 158 165	CXXCXXG MOTIF.
FT	REPEAT 184 191	CXXCXXG MOTIF.
FT	REPEAT 198 205	CXXCXXG MOTIF.
FT	REPEAT 141 141	ZINC 1 (BY SIMILARITY).
FT	METAL 144 144	ZINC 1 (BY SIMILARITY).
FT	METAL 158 158	ZINC 2 (BY SIMILARITY).
FT	METAL 161 161	ZINC 2 (BY SIMILARITY).
FT	METAL 184 184	ZINC 2 (BY SIMILARITY).
FT	METAL 187 187	ZINC 2 (BY SIMILARITY).
FT	METAL 198 198	ZINC 1 (BY SIMILARITY).
FT	METAL 201 201	ZINC 1 (BY SIMILARITY).
FT	SEQUENCE 370 AA; 40580 MW; CA02FIEDFA16ESA7 CRC64;	
QY	Query Match 44.3%; Score 42.5; DB 1; Length 370;	
DB	Best Local Similarity 32.1%; Pred. No. 11;	
DB	Matches 9; Conservative 3; Mismatches 5; Indels 11; Gaps 1;	
QY	1 KOPXGLEHD-----GINGXPKG 17	
DB	216 KVPAGIDHGGQQLRLSGQEGVNGGPAG 243	
QY	1 KOPXGLEHD-----GINGXPKG 17	
DB	216 KVPAGIDHGGQQLRLSGQEGVNGGPAG 243	

DT	16-OCT-2001 (Rel. 40, Last annotation update)	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Chaperone protein dnaJ	
GN	DNAJ OR BH1348	
OS	Bacillus halodurans	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/Staphylococcus group; Bacillus	
OX	NCBI_TaxID=86665;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C-125 / JCM 9153;	
RX	MEDLINE=20512582; PubMed=11058132;	
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,	
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,	
RA	Horikoshi K.;	
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus	
RT	halodurans and genomic sequence comparison with Bacillus subtilis."	
RL	Nucleic Acids Res. 28:4317-4331(2000).	
CC	-!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,	
CC	THE ATPASE ACTIVITY OF DNAJ (BY SIMILARITY).	
CC	-!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).	
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).	
CC	-!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	EMBL; AF045564; AAD02415.1;	
DR	InterPro: IPR004142; Ndr.	
DR	Pfam; PF03096; Ndr; 1.	
KW	Developmental protein.	
FT	DOMAIN 253 256	POLY-THR.
FT	SEQUENCE 352 AA; 38487 MW; EF35D773B9D255D9 CRC64;	
QY	Query Match 44.8%; Score 43; DB 1; Length 352;	
DB	Best Local Similarity 60.0%; Pred. No. 8.5;	
DB	Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;	
QY	3 PXGLEHGGINGXPKG 17	
DB	14 PYGLLHVIRGSPKG 28	
QY	3 PXGLEHGGINGXPKG 17	
DB	14 PYGLLHVIRGSPKG 28	

DT	16-OCT-2001 (Rel. 40, Last annotation update)	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Chaperone protein dnaJ	
GN	DNAJ OR BH1348	
OS	Bacillus halodurans	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/Staphylococcus group; Bacillus	
OX	NCBI_TaxID=86665;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C-125 / JCM 9153;	
RX	MEDLINE=20512582; PubMed=11058132;	
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,	
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,	
RA	Horikoshi K.;	
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus	
RT	halodurans and genomic sequence comparison with Bacillus subtilis."	
RL	Nucleic Acids Res. 28:4317-4331(2000).	
CC	-!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,	
CC	THE ATPASE ACTIVITY OF DNAJ (BY SIMILARITY).	
CC	-!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).	
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).	
CC	-!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	EMBL; AF045564; AAD02415.1;	
DR	InterPro: IPR004142; Ndr.	
DR	Pfam; PF03096; Ndr; 1.	
KW	Developmental protein.	
FT	DOMAIN 253 256	POLY-THR.
FT	SEQUENCE 352 AA; 38487 MW; EF35D773B9D255D9 CRC64;	
QY	Query Match 44.8%; Score 43; DB 1; Length 352;	
DB	Best Local Similarity 60.0%; Pred. No. 8.5;	
DB	Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;	
QY	3 PXGLEHGGINGXPKG 17	
DB	14 PYGLLHVIRGSPKG 28	
QY	3 PXGLEHGGINGXPKG 17	
DB	14 PYGLLHVIRGSPKG 28	

DT	16-OCT-2001 (Rel. 40, Last annotation update)	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Chaperone protein dnaJ	
GN	DNAJ OR BH1348	
OS	Bacillus halodurans	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/Staphylococcus group; Bacillus	
OX	NCBI_TaxID=86665;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C-125 / JCM 9153;	
RX	MEDLINE=20512582; PubMed=11058132;	
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,	
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,	
RA	Horikoshi K.;	
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus	
RT	halodurans and genomic sequence comparison with Bacillus subtilis."	
RL	Nucleic Acids Res. 28:4317-4331(2000).	
CC	-!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,	
CC	THE ATPASE ACTIVITY OF DNAJ (BY SIMILARITY).	
CC	-!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).	
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).	
CC	-!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	EMBL; AF045564; AAD02415.1;	
DR	InterPro: IPR004142; Ndr.	
DR	Pfam; PF03096; Ndr; 1.	
KW	Developmental protein.	
FT	DOMAIN 253 256	POLY-THR.
FT	SEQUENCE 352 AA; 38487 MW; EF35D773B9D255D9 CRC64;	
QY	Query Match 44.8%; Score 43; DB 1; Length 352;	
DB	Best Local Similarity 60.0%; Pred. No. 8.5;	
DB	Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;	
QY	3 PXGLEHGGINGXPKG 17	
DB	14 PYGLLHVIRGSPKG 28	
QY	3 PXGLEHGGINGXPKG 17	
DB	14 PYGLLHVIRGSPKG 28	

DT	16-OCT-2001 (Rel. 40, Last annotation update)	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Chaperone protein dnaJ	
GN	DNAJ OR BH1348	
OS	Bacillus halodurans	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/Staphylococcus group; Bacillus	
OX	NCBI_TaxID=86665;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C-125 / JCM 9153;	
RX	MEDLINE=20512582; PubMed=11058132;	
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,	
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,	
RA	Horikoshi K.;	
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus	
RT	halodurans and genomic sequence comparison with Bacillus subtilis."	
RL	Nucleic Acids Res. 28:4317-4331(2000).	
CC	-!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,	
CC	THE ATPASE ACTIVITY OF DNAJ (BY SIMILARITY).	
CC	-!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).	
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).	
CC	-!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	EMBL; AF045564; AAD02415.1;	
DR	InterPro: IPR004142; Ndr.	
DR	Pfam; PF03096; Ndr; 1.	
KW	Developmental protein.	
FT	DOMAIN 253 256	POLY-THR.
FT	SEQUENCE 352 AA; 38487 MW; EF35D773B9D255D9 CRC64;	
QY	Query Match 44.8%; Score 43; DB 1; Length 352;	
DB	Best Local Similarity 60.0%; Pred. No. 8.5;	
DB	Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;	
QY	3 PXGLEHGGINGXPKG 17	
DB	14 PYGLLHVIRGSPKG 28	
QY	3 PXGLEHGGINGXPKG 17	
DB	14 PYGLLHVIRGSPKG 28	

DT	16-OCT-2001 (Rel. 40, Last annotation update)	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Chaperone protein dnaJ	
GN	DNAJ OR BH1348	
OS	Bacillus halodurans	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/Staphylococcus group; Bacillus	
OX	NCBI_TaxID=86665;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C-125 / JCM 9153;	
RX	MEDLINE=20512582; PubMed=11058132;	
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,	
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,	
RA	Horikoshi K.;	
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus	
RT	halodurans and genomic sequence comparison with Bacillus subtilis."	
RL	Nucleic Acids Res. 28:4317-4331(2000).	
CC	-!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,	
CC	THE ATPASE ACTIVITY OF DNAJ (BY SIMILARITY).	
CC	-!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).	
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).	
CC	-!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	EMBL; AF045564; AAD02415.1;	
DR	InterPro: IPR004142; Ndr.	
DR	Pfam; PF03096; Ndr; 1.	
KW	Developmental protein.	
FT	DOMAIN 253 256	POLY-THR.
FT	SEQUENCE 352 AA; 38487 MW; EF35D773B9D255D9 CRC64;	
QY	Query Match 44.8%; Score 43; DB 1; Length 352;	
DB	Best Local Similarity 60.0%; Pred. No. 8.5;	
DB	Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;	
QY	3 PXGLEHGGINGXPKG 17	
DB	14 PYGLLHVIRGSPKG 28	
QY	3 PXGLEHGGINGXPKG 17	
DB	14 PYGLLHVIRGSPKG 28	

DT	16-OCT-2001 (Rel. 40, Last annotation update)	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Chaperone protein dnaJ	
GN	DNAJ OR BH1348	
OS	Bacillus halodurans	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/Staphylococcus group; Bacillus	
OX	NCBI_TaxID=86665;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C-125 / JCM 9153;	
RX	MEDLINE=20512582; PubMed=11058132;	
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,	
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,	
RA	Horikoshi K.;	
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus	
RT	halodurans and genomic sequence comparison with Bacillus subtilis."	
RL	Nucleic Acids Res. 28:4317-4331(2000).	
CC	-!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,	
CC	THE ATPASE ACTIVITY OF DNAJ (BY SIMILARITY).	
CC	-!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).	
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).	
CC	-!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	EMBL; AF045564; AAD02415.1;	
DR	InterPro: IPR004142; Ndr.	
DR	Pfam; PF03096; Ndr; 1.	
KW	Developmental protein.	
FT	DOMAIN 253 256	POLY-THR.

```

RESULT 7
YC80_GUITH          STANDARD;          PRT;      282 AA.
AC 078449;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 13-JUL-1999 (Rel. 38, Last annotation update)
DE Hypothetical 33.2 kDa protein ycf80.
GN YCF80.
OS Guillardia theta (Cryptomonas phi).
OC Chloroplast.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99128221; PubMed=9929392;
RA Douglas S.E., Penny S.L.;
RT "The plastid genome of the cryptophyte alga, Guillardia theta:
RT complete sequence and conserved syntenic groups confirm its common
RT ancestry with red algae.";
RL J. Mol. Evol. 48:236-244(1999).
CC -!- SIMILARITY: BELONGS TO THE YCF80 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF041468; AAC35638.1; -
KW Hypothetical protein; Chloroplast.
SQ SEQUENCE 282 AA; 33221 MW; 0D84447DCADA943A CRC64;

Query Match          43.8%; Score 42; DB 1; Length 282;
Best Local Similarity 58.3%; Pred. No. 9.7;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 KDPXGLEHDGIN 12
DB 175 KDPXGVNVGVN 186
|||||:|:|

RESULT 8
NDR3_MOUSE
ID NDR3_MOUSE          STANDARD;          PRT;      375 AA.
AC Q90YF9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE NDRG3 protein (Ndr3 protein).
DN NDRG3 OR NDR3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20050077; PubMed=10581191;
RA Okuda T., Kondoh H.;
RT "Identification of new genes ndr2 and ndr3 which are related to
RT Ndr1/RTP/Drg1 but show distinct tissue specificity and response to
RT N-myc.";
RL Blochem. Biophys. Res. Commun. 266:208-215(1999).
CC -!- TISSUE SPECIFICITY: THYMUS, NERVOUS SYSTEM, GUT AND KIDNEY.
CC -!- DEVELOPMENTAL STAGE: ITS EXPRESSION IS ALREADY SIGNIFICANT AT 9.5
CC DPC, COVERING THE ENTIRE EMBRYO EXCEPT THE HEART, AND IT SHOWS
CC ONLY A SLIGHT INCREASE IN LATER DEVELOPMENTAL STAGES.
CC -!- SIMILARITY: BELONGS TO THE NDRG FAMILY.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB033922; BAA85883.1; -
DR MGD; MGI:1352499; Ndr3.
DR InterPro; IPR000379; Est_lip_thioest_actsite.
DR InterPro; IPR004142; Ndr.
DR Pfam; PF03096; Ndr; 1.
SQ SEQUENCE 375 AA; 41555 MW; B33D3CC3E816AE1 CRC64;

Query Match          43.8%; Score 42; DB 1; Length 375;
Best Local Similarity 53.3%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 PXGLEHDGNGXPKG 17
DB 38 PHGMVHTVIRGLPKG 52
|||||:|:|

RESULT 9
FKH2_XENLA
ID FKH2_XENLA          STANDARD;          PRT;      427 AA.
AC P32315;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE XFKH2 protein.
GN XFKH2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8395;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94074768; PubMed=8253274;
RA Bolce M.E., Hemmati-Briylanlou A., Harland R.M.;
RT "XFKH2, a Xenopus HNF-3 alpha homologue, exhibits both
RT activin-inducible and autonomous phases of expression in early
RT embryos.";
RL Dev. Biol. 160:413-423(1993).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: PRESENT IN THE VEGETAL POLE AND MARGINAL ZONE
CC BUT NOT THE ANIMAL POLE OF GASTRULAE AND IN EQUAL LEVELS IN THE
CC DORSAL AND VENTRAL HALVES OF BOTH GASTRULAE AND NEURULAE.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M93658; AAA17050.1; -
DR HSSP; Q63245; 2HIF.
DR InterPro; IPR001766; Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00659; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein.
FT DNA_BIND 156 247
SQ SEQUENCE 427 AA; 46572 MW; 2D29A42AF960730C CRC64;

```

Tue Jul 16 16:10:14 2002

us-09-394-019a-248.rsp

```

Query Match      43.8%; Score 42; DB 1; Length 427;
Best Local Similarity 57.1%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 PXGLEHGGINGXPK 16
   | |||:| | | |
Db 301 PQALEHNGSNGEMK 314

RESULT 10
LMP2_EBV
ID LMP2_EBV STANDARD; PRT; 497 AA.
AC P13285;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Gene terminal protein (Membrane protein LMP-2A/LMP-2B).
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88283646; PubMed=2840285;
RA Laux G., Perricaudet M., Farrell P.J.;
RT "A spliced Epstein-Barr virus gene expressed in immortalized
   lymphocytes is created by circularization of the linear viral
   genome.";
RL EMO J. 7:769-774(1988).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89095024; PubMed=2536113;
RA Sample J., Liebowitz D., Kieff E.;
RT "Two related Epstein-Barr virus membrane proteins are encoded by
   separate genes.";
RL J. Virol. 63:933-937(1989).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -!- FUNCTION: MAY BE IMPORTANT FOR IMMORTALIZATION OF CELLS BY EBV.
CC -!- SUBCELLULAR LOCATION: MEMBRANE (PROBABLE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24212; AAA45887.1; -
CC EMBL; Y00835; CAA68762.1; -
CC EMBL; V01555; -; NOT_ANNOTATED_CDS.
CC PIR; A30178; WNBELM.
KW Transmembrane.
FT CHAIN 1 497 MEMBRANE PROTEIN LMP-2A.
FT CHAIN 120 497 MEMBRANE PROTEIN LMP-2B.
FT TRANSMEM 122 141 POTENTIAL.
FT TRANSMEM 150 168 POTENTIAL.
FT TRANSMEM 178 198 POTENTIAL.
FT TRANSMEM 208 235 POTENTIAL.
FT TRANSMEM 242 259 POTENTIAL.
FT TRANSMEM 267 288 POTENTIAL.
FT TRANSMEM 300 316 POTENTIAL.
FT TRANSMEM 321 339 POTENTIAL.
FT TRANSMEM 355 373 POTENTIAL.

Query Match      43.8%; Score 42; DB 1; Length 497;
Best Local Similarity 63.6%; Pred. No. 18;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 GLEHGGINGXP 15
   |||:| | | |
Db 87 GLQHDGNDGLP 97

RESULT 11
TAP1_MOUSE
ID TAP1_MOUSE STANDARD; PRT; 724 AA.
AC P21958;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigen peptide transporter 1 (APT1) (Histocompatibility antigen
   modifier 1).
DE ABCB2 OR TAP1 OR HAM-1.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=WEHI-3;
RX MEDLINE=97307601; PubMed=9164943;
RA Marusina K., Iyer M., Monaco J.J.;
RT "Allelic variation in the mouse Tap-1 and Tap-2 transporter genes.";
RL J. Immunol. 158:5251-5256(1997).
[2]
RP SEQUENCE OF 148-724 FROM N.A.
RX MEDLINE=91102550; PubMed=2270487;
RA Monaco J.J., Cho S., Attaya M.;
RT "Transport protein genes in the murine MHC: possible implications for
   antigen processing.";
RL Science 250:1723-1726(1990).
[3]
RP SEQUENCE OF 148-724 FROM N.A.
RX STRAIN=BALB/C;
RA Sun X.Y., Zhou J., Frazer I.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE TRANSPORT OF ANTIGENS FROM THE CYTOPLASM
CC TO A MEMBRANE-BOUND COMPARTMENT FOR ASSOCIATION WITH MHC CLASS I
CC MOLECULES.
CC -!- SUBUNIT: HETERODIMER OF TAP1 AND TAP2.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U60018; AAA11962.1; -
CC EMBL; M55637; AAA39570.1; -
CC EMBL; X59615; CAA42178.1; -
CC PIR; A37779; A37779.
CC MGD; MGI:98483; Abcb2.
CC InterPro; IPR003593; AAA.
CC InterPro; IPR001140; ABC_transporter_tmem.
CC InterPro; IPR003439; ABC_transportr.
CC InterPro; IPR001687; ATP_GTP_A.
CC Pfam; PF00664; ABC_membrane; 1.

```

RESULT	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100											
RESULT	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100											
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	3																																																																				

```

or send an email to license@isb-sib.ch).
-----
CC EMBL; AL627265; CAD01321.1;
CC DR Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
CC KW SEQUENCE 327 AA; 35041 MW; F7F563DB5326FEB CRC64;
SQ
Query Match 43.2%; Score 41.5; DB 1; Length 327;
Best Local Similarity 39.1%; Pred. No. 14; Gaps 1;
Matches 9; Conservative 1; Mismatches 4; Indels 9; Gaps 1;
QY 2 DPXG-----LEHGDINGXP 15
| | | | |
DB 12 DPAGIGPEIIIVKALSEDGLNGAP 34
RESULT 14
PX2_SALTY STANDARD; PRT; 327 AA.
ID PXA2_SALTY
AC P58718;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-WAR-2002 (Rel. 41, Last sequence update)
DT 01-WAR-2002 (Rel. 41, Last annotation update)
DE 4-hydroxythreonine-4-phosphate dehydrogenase 2 (EC 1.1.1.262) (4-
GN (phosphohydroxy)-L-threonine dehydrogenase 2).
DE PDXA2 OR STM0163.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -1- FUNCTION: Catalyzes the NAD-dependent oxidation of 4-
CC (phosphohydroxy)-L-threonine (HTP) into 2-amino-3-oxo-4-
CC (phosphohydroxy)butyric acid which spontaneously decarboxylate to
CC form 1-amino-3-(phosphohydroxy)propan-2-one (3-amino-2-oxopropyl
CC phosphate) (By similarity).
CC -1- CATALYTIC ACTIVITY: 4-(phosphonoxy)-threonine + NAD(+) = 2-
CC amino-3-oxo-4-phosphonooxobuturate + NADH.
CC -1- PATHWAY: De novo synthesis of pyridoxine (vitamin B6) and
CC pyridoxal phosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE PDXA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; AE008701; AAL19127.1;
DR Stygene; SG77777; pdxa2.
KW Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
SQ SEQUENCE 327 AA; 35064 MW; 03BB6725F1896440 CRC64;
Query Match 43.2%; Score 41.5; DB 1; Length 327;
Best Local Similarity 39.1%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 4; Indels 9; Gaps 1;

```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M31162; AAA28859.1; ..  
DR EMBL; AE003615; AAF52471.1; ..  
DR PIR; A36311; A36311.  
DR HSP; P09651; IHA1.  
DR FlyBase; FBgn0016978; snRNP70K.  
DR InterPro; IP8000504; RRM.  
DR Pfam; PF00076; rrm; 1.  
DR SMART; SM00360; RRM; 1.  
DR PROSITE; PS00102; RRM; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
KW Nuclear protein; Ribonucleoprotein; RNA-binding; mRNA processing.  
FT DOMAIN 102 180 RNA-BINDING (RRM).  
FT DOMAIN 254 350 ARG/GLU-RICH (MIXED CHARGE).  
FT CONFLICT 278 278 N -> S (IN REF. 1).  
SQ SEQUENCE 448 AA; 52900 MW; 0DDFB5A39CA72AEB CRC64;

Query Match 42.7%; Score 41; DB 1; Length 448;  
Best Local Similarity 61.5%; Pred. No. 24;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 6 LEHDSGXPKGY 18  
| | | | |  
Db 133 LIHQESGKPKGY 145

Search completed: July 16, 2002, 11:22:17  
Job time: 115 sec

**THIS PAGE BLANK (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2002, 11:16:57 ; Search time 25.08 seconds  
(without alignments)  
124.159 Million cell updates/sec

Title: US-09-394-019a-248  
Perfect score: 96  
Sequence: 1 KDPXGLEHNGXPKGY 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 08  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_virus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	50.0	590	2 Q93CA0	Q93ca0 bifidobacte
2	47	49.0	451	5 Q9VPG0	Q9vpg0 drosophila
3	45	46.9	135	16 Q9A3D9	Q9a3d9 caulobacter
4	44.5	46.4	117	12 Q69133	Q69133 human herpe
5	44.5	46.4	118	12 Q66567	Q66567 human herpe
6	44.5	46.4	118	12 Q66568	Q66568 human herpe
7	44.5	46.4	118	12 Q66570	Q66570 human herpe
8	44.5	46.4	118	12 Q66573	Q66573 human herpe
9	44.5	46.4	118	12 Q66550	Q66550 human herpe
10	44.5	46.4	118	12 Q66551	Q66551 human herpe
11	44.5	46.4	118	12 Q66552	Q66552 human herpe
12	44.5	46.4	118	12 Q66553	Q66553 human herpe
13	44.5	46.4	118	12 Q66555	Q66555 human herpe
14	44.5	46.4	118	12 Q66556	Q66556 human herpe
15	44.5	46.4	118	12 Q66557	Q66557 human herpe
16	44.5	46.4	118	12 Q66559	Q66559 human herpe

17	44.5	46.4	118	12 Q66560	Q66560 human herpe
18	44.5	46.4	118	12 Q66561	Q66561 human herpe
19	44.5	46.4	118	12 Q66562	Q66562 human herpe
20	44.5	46.4	118	12 Q66563	Q66563 human herpe
21	44.5	46.4	118	12 Q66565	Q66565 human herpe
22	44.5	46.4	183	12 Q69134	Q69134 human herpe
23	44.5	46.4	542	16 Q9X287	Q9x287 thermotoga
24	44.5	46.4	826	5 Q9B130	Q9b130 ciona intes
25	43.5	45.3	655	4 Q96TA8	Q96ta8 homo sapien
26	43.5	45.3	655	4 Q96LDB	Q96ldb homo sapien
27	43.5	45.3	1146	2 Q68317	Q68317 vibrio chol
28	43.5	45.3	1146	16 Q9KRI8	Q9kri8 vibrio chol
29	43	44.8	339	4 Q9GZM3	Q9gzm3 homo sapien
30	43	44.8	339	4 Q96PL9	Q96pl9 homo sapien
31	43	44.8	339	11 Q923D7	Q923d7 mus musculu
32	43	44.8	352	4 Q9GZX0	Q9gzx0 homo sapien
33	43	44.8	371	4 Q9GZM1	Q9gzm1 homo sapien
34	42.5	44.3	183	5 Q9VJ96	Q9vj96 drosophila
35	42	43.8	73	6 Q28751	Q28751 ov18 sp. ke
36	42	43.8	118	12 Q66566	Q66566 human herpe
37	42	43.8	118	12 Q66569	Q66569 human herpe
38	42	43.8	118	12 Q66571	Q66571 human herpe
39	42	43.8	118	12 Q66572	Q66572 human herpe
40	42	43.8	118	12 Q66574	Q66574 human herpe
41	42	43.8	118	12 Q66554	Q66554 human herpe
42	42	43.8	118	12 Q66558	Q66558 human herpe
43	42	43.8	129	11 Q63911	Q63911 mus sp. tap
44	42	43.8	200	16 Q9KU09	Q9kuq9 vibrio chol
45	42	43.8	211	4 Q9BY23	Q9by23 homo sapien

ALIGNMENTS

RESULT 1

Q93CA0 PRELIMINARY; PRT; 590 AA.  
AC Q93CA0;  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE ALPHEA-GLUCOSIDASE.  
GN AGLB.  
OS Bifidobacterium adolescentis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.  
OX NCBI\_TaxID=1680;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 20083.  
RA Van den Broek L.A.M., Struijs K., Voragen A.G.J., Verdoes J.C.,  
RA Belman G.;  
RT "Cloning and characterization of two alpha-glucosidases from  
RT Bifidobacterium adolescentis.";  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF411186; AAL05573.1;  
SQ SEQUENCE 590 AA; 66575 MW; 545933DDC93B3CEB CRC64;

Query Match

Best Local Similarity 50.0%; Score 48; DB 2; Length 590;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 KDPXGLEHNGXPKG 17

Db 521 EDQPGKANDGANGPFGG 537

RESULT 2

Q9VPG0 PRELIMINARY; PRT; 451 AA.  
ID Q9VPG0;  
AC Q9VPG0;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)

Tue Jul 16 16:10:14 2002

```

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CG5282 PROTEIN.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY; PubMed=10731132;
RX MEDLINE=20196006; Pubmed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya P., Brothier P.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Cantor A., Chandra I.,
RA Burdick K.C., Busam D.A., Butler J., Cahoon E., Cantor A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fouts D.C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hartzell N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA "the genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003591; AAF51593.1; -.
DR MEROPS; M19.UNW; -.
DR FlyBase; FBgn0036986; CG5282.
DR InterPro; IPR000180; Renal_dipeptidase.
DR Pfam; PF01244; Renal_dipeptidase; 1.
SQ SEQUENCE 451 AA; 49262 MW; 95103A0674F25C0F CRC64;

Query Match 49.08; Score 47; DB 5; Length 451;
Best Local Similarity 50.0%; Pred. No 9.8;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 GLEHGGINGXPKGY 18
DB 375 GVDHIGLGAPKSY 388

RESULT 3
ID Q9A3D9 PRELIMINARY; PRT; 135 AA.
AC Q9A3D9;
DR 01-JUN-2001 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN CC3265.
GN Caulobacter crescentus.
OS Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter
OC Caulobacter
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; Pubmed=11259647;
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uitterback T., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005990; AAK25227.1; -.
DR TIGR; CC3265; -.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 135 AA; 14607 MW; 01C21B2C6B498805 CRC64;

Query Match 46.9%; Score 45; DB 16; Length 135;
Best Local Similarity 50.0%; Pred. No 5.7;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 GLEHGGINGXPKGY 18
DB 89 GMDFDGKGLPKGW 102

RESULT 4
ID Q69133 PRELIMINARY; PRT; 117 AA.
AC Q69133;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE LATENT MEMBRANE PROTEIN 2A (FRAGMENT).
GN LMP 2A.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88283646; Pubmed=2840285;
RA Laux G., Perricaudet M., Farrell P.J.;
RT "A spliced Epstein-Barr virus gene expressed in immortalized
RT lymphocytes is created by circularization of the linear viral
RT genome."
RL EMBO J. 7:769-774(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89095024; Pubmed=2536113;
RA Sample J., Liebowitz D., Kieff E.D.;
RT "Two related Epstein-Barr virus membrane proteins are encoded by
RT separate genes."
RL J. Virol. 63:933-937(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219426; Pubmed=1313931;
RA Busson P., McCoy R., Sadler R., Gilligan K., Tursz T., Raab-Traub N.;
RT "Consistent transcription of the Epstein-Barr virus LMP2 gene in
RT nasopharyngeal carcinoma."
RL J. Virol. 66:3257-3262(1992).
DR EMBL; M87778; AAA45885.1; -.
DR NON_TER 1

```

Query Match 46.4%; Score 44.5; DB 12; Length 118;

RESULT	8
Q66573	
ID	PRELIMINARY; PRT; 118 AA.
AC	Q66573;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-DIC-2001 (TrEMBLrel. 19, Last annotation update)
DE	LATENT MEMBRANE PROTEIN 2A (FRAGMENT).
GN	LMP2.
OS	Human herpesvirus 4 (Epstein-Barr virus).
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
CC	Gammapherpesvirinae; Lymphocryptovirus.
OX	NCBI_TaxID=10376;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAINE-BL72 BURKITT LYMPHOMA FROM ALGERIA;
RX	MEDLINE=95145941; PubMed=7844523;
RA	Bussón P., Hood Edwards R., Tursz T., Raab-Traub N. ;
RT	"Sequence polymorphism in the Espdtein-Barr virus latent membrane protein (lmp)-2 gene.";
RL	J. Gen. Virol. 76:139-145(1995).
FT	EMBL: X81763; CAA57367.1; -.
DR	-
DT	NON_TER 118
FT	SEQUENCE 118 AA; 12445 MW; C73A8577EA15ADEB CRC64;
SQ	

Query Match	46.4%	Score 44.5;	DB 12;	Length 118;
Best Local Similarity	50.0%	Prod. No. 5.9;		
Matches	9;	Conservative	3;	Mismatches 3;
				Indels 3

QY	1	KDPX---GLEHGDINGXP 15 ::: 
Db	80	QDPSLYLGLQHNDGNDGLP 97
RESULT	9	
ID Q66550	PRELIMINARY;	PRT; 118 AA.
DT 01-NOV-1996	(TREMELrel. 01, Created)	
DT 01-NOV-1996	(TREMELrel. 01, Last sequence update)	
DT 01-DEC-2001	(TREMELrel. 19, Last annotation update)	
DE LATENT MEMBRANE PROTEIN 2A (FRAGMENT).		
GN LMP2.		
OS Human herpesvirus 4 (Epstein-Barr virus).		
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
OC Gammaherpesvirinae; Lymphocryptovirus.		
OX NCBI_TaxID=10376;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC SPRAIN-D6 NASOPHARYNGEAL CARCINOMA BIOPSY SAMPLE FROM CHINA;		
RX MEDLINE=95146941; PubMed=7844523;		
RA Busson P., Hood Edwards R., Tursz T., Raab-Traub N.;		
RT "Sequence polymorphism in the Epstein-Barr virus latent membrane protein (lmp)-2 gene.";		
RL J. Gen. Virol. 76:139-145(1995).		
RL EMBL; X81766; CAA57370.1; -.		
FT NON_TER 118		
FT SEQUENCE 118 AA; 12507 MW; OA76C05F48840069 CRC64;		
SQ		
Query Match 46.4%; Score 44.5; DB 12; Length 118; Best Local Similarity 50.0%; Pred. No. 5.9; Indels 3; Gaps 1; Matches 9; Conservative 3; Mismatches 3; Indels 3; Gaps 1;		
QY	1	KDPX---GLEHGDINGXP 15 ::: 
Db	80	QDPSLYLGLQHNDGNDGLP 97
RESULT	10	
ID Q66551	PRELIMINARY;	PRT; 118 AA.
AC Q66551;		
DT 01-NOV-1996	(TREMELrel. 01, Created)	
DT 01-NOV-1996	(TREMELrel. 01, Last sequence update)	
DT 01-DEC-2001	(TREMELrel. 19, Last annotation update)	
DE LATENT MEMBRANE PROTEIN 2A (FRAGMENT).		
GN LMP2.		
OS Human herpesvirus 4 (Epstein-Barr virus).		
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
OC Gammaherpesvirinae; Lymphocryptovirus.		
OX NCBI_TaxID=10376;		
RN [1]		
RP SEQUENCE FROM N.A.		
RX MEDLINE=95146941; PubMed=7844523;		
RA Busson P., Hood Edwards R., Tursz T., Raab-Traub N.;		
RT "sequence polymorphism in the Epstein-Barr virus latent membrane protein (lmp)-2 gene.";		
RL J. Gen. Virol. 76:139-145(1995).		
RL EMBL; X81767; CAA57371.1; -.		
FT NON_TER 118		
FT SEQUENCE 118 AA; 12461 MW; AB3A996BEA15ADFE CRC64;		
SQ		
Query Match 46.4%; Score 44.5; DB 12; Length 118; Best Local Similarity 50.0%; Pred. No. 5.9; Indels 3; Gaps 1; Matches 9; Conservative 3; Mismatches 3; Indels 3; Gaps 1;		
QY	1	KDPX---GLEHGDINGXP 15 ::: 
Db	80	QDPSLYLGLQHNDGNDGLP 97

DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE LATENT MEMBRANE PROTEIN 2A (FRAGMENT).  
GN LMP2.  
OS Human herpesvirus 4 (Epstein-Barr virus).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Lymphocryptovirus.  
OX NCBI\_TaxID=10376;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JIOYE AFRICAN BURKITT LYMPHOMA;  
RX MEDLINE=95146941; PubMed=7844523;  
RA Busson P., Hood Edwards R., Tursz T., Raab-Traub N.;  
RT "Sequence polymorphism in the Espdtein-Barr virus latent membrane  
protein (lmp)-2 gene."  
RL J. Gen. Virol. 76:139-145(1995).  
DR EMBL; X81771; CAA57375.1; -  
FT NON\_TER 118 118  
SQ SEQUENCE 118 AA; 12539 MW; 7B34420DE0B51373 CRC64;

Query Match 46.4%; Score 44.5; DB 12; Length 118;  
Best Local Similarity 50.0%; Pred. No. 5.9;  
Matches 9; Conservative 3; Mismatches 3; Indels 3; Gaps 1;  
Qy 1 KDPX---GLEHGDINGXP 15  
:|:|:|:|:|:|  
Db 80 QDPSLYLGLQHDGNDGLP 97

RESULT 14  
O66556  
ID O66556 PRELIMINARY; PRT; 118 AA.  
AC O66556;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE LATENT MEMBRANE PROTEIN 2A (FRAGMENT).  
GN LMP2.  
OS Human herpesvirus 4 (Epstein-Barr virus).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Lymphocryptovirus.  
OX NCBI\_TaxID=10376;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L2 NASOPHARYNGEAL CARCINOMA BIOPSY SAMPLE FROM CHINA;  
RX MEDLINE=95146941; PubMed=7844523;  
RA Busson P., Hood Edwards R., Tursz T., Raab-Traub N.;  
RT "Sequence polymorphism in the Espdtein-Barr virus latent membrane  
protein (lmp)-2 gene."  
RL J. Gen. Virol. 76:139-145(1995).  
DR EMBL; X81772; CAA57376.1; -  
FT NON\_TER 118 118  
SQ SEQUENCE 118 AA; 12461 MW; AB3A996BEA15ADFE CRC64;

Query Match 46.4%; Score 44.5; DB 12; Length 118;  
Best Local Similarity 50.0%; Pred. No. 5.9;  
Matches 9; Conservative 3; Mismatches 3; Indels 3; Gaps 1;  
Qy 1 KDPX---GLEHGDINGXP 15  
:|:|:|:|:|:|  
Db 80 QDPSLYLGLQHDGNDGLP 97

RESULT 15  
O66557  
ID O66557 PRELIMINARY; PRT; 118 AA.  
AC O66557;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE LATENT MEMBRANE PROTEIN 2A (FRAGMENT).  
GN LMP2.  
OS Human herpesvirus 4 (Epstein-Barr virus).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Lymphocryptovirus.  
OX NCBI\_TaxID=10376;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JIOYE BURKITT LYMPHOMA FROM AFRICA;  
RX MEDLINE=95146941; PubMed=7844523;  
RA Busson P., Hood Edwards R., Tursz T., Raab-Traub N.;  
RT "Sequence polymorphism in the Espdtein-Barr virus latent membrane  
protein (lmp)-2 gene."  
RL J. Gen. Virol. 76:139-145(1995).  
DR EMBL; X81773; CAA57377.1; -  
FT NON\_TER 118 118  
SQ SEQUENCE 118 AA; 12480 MW; 9BD44A77EA15B979 CRC64;

Query Match 46.4%; Score 44.5; DB 12; Length 118;  
Best Local Similarity 50.0%; Pred. No. 5.9;  
Matches 9; Conservative 3; Mismatches 3; Indels 3; Gaps 1;  
Qy 1 KDPX---GLEHGDINGXP 15  
:|:|:|:|:|:|  
Db 80 QDPSLYLGLQHDGNDGLP 97

Search completed: July 16, 2002, 11:21:59  
Job time: 302 sec

us-09-394-019a-248.rspt

Tue Jul 16 16:10:14 2002

**THIS PAGE BLANK (rspt)**